“The Soul is the Prison of the Body” – Mandatory Moral Enhancement, Punishment & Rights Against Neuro-Rehabilitation

Preprint · January 2018

CITATIONS
0

READS
103

1 author:

Christoph Bublitz
University of Hamburg
35 PUBLICATIONS **360** CITATIONS

SEE PROFILE

Forthcoming in: David Birks and Thomas Douglas, eds. *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice*. Oxford: Oxford University Press, 2017.

"The Soul is the Prison of the Body" – Mandatory Moral Enhancement, Punishment & Rights against Neuro-Rehabilitation

Abstract

The promise of neurobiological interventions that afford improving pro-social behavior is particularly interesting for criminal justice systems. After all, rehabilitation of offenders is one of their central objectives. This raises the question whether states can deploy such means to rehabilitate offenders against the latters' will, as part of – or instead of – punishment. Some advocates of compulsory treatments of offenders consider them more humane (and effective) than current forms of hard treatment such as incarceration. This chapter critically engages with suggestions to treat legally competent offenders for rehabilitative purposes against their will by emphasizing two aspects: First, strong human rights of offenders – summarily the right to mental self-determination – oppose mandatory interventions into criminogenic psychological states or processes. These human rights are not (yet) recognized in every jurisdiction, but emerge from general liberal and democratic principles most western jurisdictions endorse. Secondly, the case for mandatory rehabilitation is weaker than it may appear at first glance because it is anything but clear that and why the penological aim of rehabilitation justifies severe interferences of offenders' rights. In any case, it seems that states could attain their legitimate forward-looking aims – preventing recidivism – by less restrictive means such incapacitation. Thus, compulsory rehabilitation may only be justified in exceptional cases. Rather, offenders should be offered a choice between neuro-rehabilitation and detention.

In dealing with criminals, there shall be regularly adopted measures of corrective study classes, individual interviews, study of assigned documents, and organized discussion, to educate them in the admission of guilt and obedience to the law, political and current events, labour production, and culture, so as to expose the nature of the crime committed, thoroughly wipe out criminal thoughts and establish a new moral code.

Prison Regulations in Maoist China, 1950s[1]

Neurotechnological interventions into minds and brains which potentially empower persons to transform themselves for the better are a welcome development, even more if they prove capable of supporting offenders to lead a law-abiding life in the future. The criminal justice system is not only well-advised, but may have moral and legal duties to offer such means to offenders, preferably alongside broader rehabilitative programs involving psychological counselling, social support and, if necessary, moral education.[2] New Retributivism, which domineered penal thought after moral reform theories had fallen out of favour in the latter half of the last century, has come under criticism from empirical and normative directions. Normatively, the elusive notion of "just desert", central to modern retributive theories, has been cast into doubt by the resurgence of philosophical objections to free will, supported by advances in neuroscience and psychology. Empirically, many Western countries grapple with high rates and costs of incarceration. Given the poor track record of prisons, recidivism rates, and the strikingly meaningless waste of lifetime to which imprisonment amounts for many offender, alternative sanctions and novel rehabilitative approaches appear highly attractive. The promise of effective methods overcoming the – in any case, erroneous – "nothing works" credo[3] may therefore reignite the controversy over the proper role of rehabilitation in the criminal justice system in general, and the legitimacy of *mandatory neurorehabilitation* in particular.[4] Arguments in favour of rehabilitating unwilling offenders without consent have been forwarded by scholars, and politicians and the public will follow suit as soon as effective neurotools become available.[5]

[1] Quoted in R. J. Lifton, 'Thought reform and the psychology of totalism: a study of brainwashing in China' (New York: Norton, 1963), p. 17. The title alludes to a line by Michel Foucault in 'Discipline and Punish', 2nd ed. (New York: Random House, 1995), p. 30.

[2] The European Court of Human Rights repeatedly held that all prisoners have to be offered the possibility of rehabilitation. Recently in *Vinter v. UK* (Jul 9th 2013–66069/09) at 114.

[3] Robert Martinson, 'New Findings, New Views: A Note of Caution Regarding Sentencing Reform', Hofstra Law Review (1979), 243–258; Cullen/Gilbert, 'Reaffirming Rehabilitation' 2nd ed. (Amsterdam: Elsevier, 2013); Even psychopathy might be treatable, Kiehl/Hoffman, 'The Criminal Psychopath', Jurimetrics 51 (2011), 355–397.

[4] Greely, 'Neuroscience and Criminal Justice: Not Responsibility but Treatment' Kansas Law Review 56 (2008) 1103; Thomas Douglas, 'Moral Enhancement', Journal of Applied Philosophy 25 (2008), 228-245. N.A. Vincent, 'Restoring responsibility: Promoting justice, therapy and reform through direct brain interventions', Criminal Law and Philosophy 8 (2014), 21-42; Elizabeth Shaw, 'Direct brain interventions and responsibility enhancement', Criminal Law and Philosophy 8 (2014), 1-20; Walter Glannon, 'Intervening in the psychopaths brain', Theoretical Medicine Bioethics 35 (2014), 43-57.

[5] Adrian Raine, 'The Anatomy of Violence' (London: Allen Lane, 2013); Thomas Douglas, 'Criminal Rehabilitation Through Medical Intervention: Moral Liability and the Right to Bodily Integrity', Journal of Ethics 18 (2014), 101-122.

Without rehearsing the historical debate over rehabilitation (which was deeply entangled with abuses of therapeutic powers and, in some jurisdictions, indeterminate sentencing),[6] this chapter addresses the question whether offenders who refuse to participate can be subjected to neurorehabilitative measures against their will, as *part of their punishment*.[7] Reasons for refusing rehabilitation can be manifold, from rejecting invasive state authority over oneself to positive identification with one's supposedly immoral character traits. For the law, motives for refusal are largely irrelevant as long as offenders are competent. The necessary conditions and problems of *voluntary* participation in such programmes are not addressed in this chapter. I only wish to note that, without further reasons to the contrary, offenders' consent should not be considered invalid just because it is incentivised, coupled with benefits such as early release or parole, or offered as an alternative to incarceration.[8]

My argument against mandatory rehabilitation has three parts. To begin, I shall clarify what is in need of justification by sketching the contours of the law as it stands, and some premises of liberal democratic orders (I). The second part introduces a set of human rights of offenders that emerges from theses premises and that opposes mandatory neurorehabilitation. I summarily call it a human right to mental self-determination. Neither the law, nor philosophy, have given it sufficient attention. Any argument in favour of mandatory rehabilitation has to show the way in which it outweighs this right (II). The third part (III) critically engages with such arguments for, and potential justifications of, mandatory neurorehabilitation. I shall take objection to common, two-tiered justifications which note affirmatively that rehabilitation is an accepted penological goal, and then demonstrate that it is more humane than hard-treatment because it inflicts less suffering. However, a closer look reveals that such a line of argument is insufficient and incomplete because goals or aims, or interests and benefits, of the state, by themselves, cannot override offenders' countervailing *right* to mental self-determination. Only rights can override rights. The argument thus falls short of explaining where the right to intervene into the minds of offenders comes from (over and above the state's interests in the moral improvement of offenders). I argue that it does not arise from the punitive powers of the state in light of retributive or communicative theories. In addition, with respect to forward-looking considerations, states have a right (and duty) to prevent future offences but can regularly achieve such preventive aims through less restrictive means, mainly incapacitation. As a consequence, justifications for mandatory

[6] See e.g. C.S. Lewis, 'The Humanitarian Theory of Punishment', Res Judicatae 6 (1953), 224; F.A. Allen, The decline of the rehabilitative ideal: Penal policy and social purpose (New Haven: Yale University Press, 1981).

[7] Some may object that the concept of punishment implies suffering or hard treatment over and above therapy, so that neurorehabilitation does not qualify. I am not interested in the conceptual claim, only in whether the state can impose mandatory treatment as part of the negative sanctions upon conviction of a crime.

[8] The degree to which convicts, especially inmates, can consent to rehabilitative treatments is controversial. Some argue that consent is invalid because of the structural pressures of imprisonment; or that offers of rehabilitation instead of punishment are coercive or exploitative (e.g., Hübner/White, AJOB Neuroscience 7 (2016), 140-149. However, such criticisms seem to misunderstand voluntariness more as a mental rather than a normative concept. Although prisoners may consider both options unattractive, they are still free in a minimal but sufficient sense to reject rehabilitation. Even more: if states fail to offer methods to shorten incarceration, they are flouting their duty to interfere with liberties of offenders in the least restrictive way.

neurorehabilitation are weaker, whereas the rights of offenders against neurorehabilitation are stronger than often suggested. Therefore, not much room for mandatory neurorehabilitation *as part of penal punishment* remains. However, as the public and politicians are not particularly known for their respect for prisoners' rights, I shall conclude by sketching two lines of argument buttressing the significance of the right to mental self-determination (IV).

The range of potential novel rehabilitative methods is broad, from the nutrimental enrichment of diets, neurofeedback gaming, to pharmaceutically augmented psychotherapy, or novel brain implants. The methods of interests in this chapter are direct brain interventions, i.e. those that work mainly at the neurophysiological level, such as psychoactive drugs and brain stimulation devices, and that target mental properties such as emotions, moods, beliefs, thoughts or behavioural dispositions, with the aim of preventing future (re-)offending. Ultimately, every method has to be evaluated on its own in light of its effectiveness and side effects. But apart from these contingent empirical matters, which I cannot address here, these methods share the distinctive feature of transforming relevant risk factors of offenders at the bio-psychological level. I suggest that they differ from accepted rehabilitative ones such as education, anger management, or cognitive-behavioural training in normative terms for two reasons: such non-direct interventions are less worrisome because they engage with offenders through reason and on a different level which is, at least in principle, respectful of offenders as autonomous, self-controlling subjects and does not bypass mental control. Moreover, these interventions are likely inept at significantly transforming the mental world of addresses without their active participation. Therefore, my argument against neurorehabilitation may not fully apply to ordinary methods, although it does not presuppose a categorical distinction between these means.

Finally, to avoid misunderstandings, some words about the level of argument are in order. The following is a *legal* argument but not in the sense of a descriptive claim about what the laws in one or another jurisdiction currently allow or forbid. Instead, it is a more abstract argument based on liberal legal principles and human rights, not about how the law is but how it *should be*, according to principles it already endorses. The argument is reformative in nature as it suggests developing a stronger legal protection of the human mind. It is not confined to a specific jurisdiction; in fact, I hope it can be translated into the more specific doctrines and technical lingo of particular jurisdictions without losing its argumentative thrust. Although neither the law as it stands, nor case precedents, possess authority in this kind of argument, it is worth briefly reviewing jurisprudence for several reasons. The law, for instance, often draws helpful distinctions between case-types and justifications which also apply to, but are not regularly found in the same complexity in purely philosophical arguments especially those employing a case-based reasoning methodology. Secondly, one can conceive of the law as a systematic set of interrelated norms that strives for internal consistency. Legal answers to specific problems have often stood the test of consistency with further premises, and are all-things-considered judgments that seek to strike just balances between various considerations. More

generally, I suggest that exchanges between philosophy and the law would benefit by taking note of those topics and arguments which are discussed in a lively manner in the other discipline, and others about which there might be ample room (and need) for interdisciplinary collaboration. Philosophers only interested in the substantive argument are invited to skip the following section (and turn to II).

Lastly, I wish to note that such abstract legal arguments, detached from concrete cases, usually do not allow for universal conclusions about the (im-)permissibility of particular practices across various contexts in each condition. Such "absolute" arguments are hard to find in the law (torture might be an exception). Regularly, the law develops a more fine-grained taxonomywhich reconciles countervailing rights rather than giving one strict priority over the other. The result of abstract legal discussions such as the present one is, thus, to stake out relevant arguments and positions which future assessments of concrete cases and specific means as well as policy decisions have to accommodate. I suggest that arguments against mandatory rehabilitation, once properly recognised, outweigh those in its favour. But this does not imply that, under extraordinary circumstances (which abound in thought experiments), the scales of justice may not also turn to the other side.

I. Liberal Premises & Legal Precedents

Whether a particular sanction is a justifiable response to a criminal offence depends on a range of questions: Is punishment legitimate? For which aims, and does attaining these aims outweigh opposing rights of offenders? Do the reasons justifying punishment, if any, also extend to particular forms? With respect to neurorehabilitation, the key question is whether the state acquires permission to alter the moral properties of a person in virtue of her having committed an offence. This corresponds, from the offender's view, to the question whether her claims against being subjected to rehabilitative measures become overridable by punitive powers of the state. I shall address these points in light of the following contours of liberal constitutional states:

In liberal legal orders, citizens enjoy wide freedoms with respect to their external conduct and even more in personal matters. Conversely, governmental powers are limited in principled ways.[9] A founding idea of liberal orders, traceable to progenitors as *Locke*, *Kant* and *Mill*, is that every person has a right to herself. A person's character, including moral beliefs and dispositions, is a self-regarding matter over which governments should have no say. The same is true for her preferences and choices, in the absence of severe decision-making deficits. The inner and outer freedoms of citizens correlate with a duty to obey the law. Liberal legal orders presume, for the sake of these freedoms, that citizens are able and willing to adjust their conduct in conformity to the law. However, while citizens owe compliance in conduct, they owe neither psychological acceptance nor affirmation of the substantive

[9] "State" and "governments" are used interchangeably, as philosophical traditions and jurisdictions seem to vary with respect to who needs to justify punitive sanctions.

content of the law. They may adhere to very different moral teachings and attempt to change laws of which they disapprove. Thus, liberal orders impose on citizens duties to behave in specific ways but leave them the mental freedom to think or feel as they please and, in particular, to disagree in moral matters. Inner and outer freedoms further correlate with citizens' responsibility for their failures to comply with the law. Breaches may incur civil and criminal liability and can, depending on circumstances, rebut the presumption of law-abidingness which may expose them to preventive measures. Apart from that, liberal orders oblige offenders neither to feel remorse or guilt for their deeds, nor to enhance their moral character. Governments may promote but cannot demand such socially desirable transformations. These liberal ideas find their expression in human rights, explored in the following section. Still, different worlds are conceivable, and so are different political orders which replace the liberal and predominantly punitive with a less liberal and preventive system. This would come with costs to liberty most of us, I suspect, are unwilling to pay. Be this as it may. The following argument is set against the backdrop of these liberal premises.

Legal Precedents

In spite of these premises, the idea of rehabilitating offenders against their will through neuro- (or as they were earlier called, psycho-)technical means is anything but new. Historically, states have subjected offenders to lobotomies and other forms of psychosurgery, electroshocks, aversive conditioning, reinforcement learning, psycho- and cognitive behavioural therapies, Antabuse and drug withdrawals, castration and psychotropic substances. Many of these rather intrusive means were either abandoned due to medical concerns (side effects, lack of efficiency), or their coercive imposition was banned for legal and political reasons. Still, most states routinely drug offenders forcibly, mainly dangerous persons who suffer from mental disorders or substance addiction. Importantly, justifications for these interventions differ from those relevant for present concerns. To appreciate the differences, some distinctions are helpful. The clearest (though not unproblematic) class of cases concerns involuntary treatment of *incompetent* mentally disordered persons for their own benefit under a classic paternalistic rationale (*parens patriae* power). Such medications are legally permissible in every jurisdiction under further conditions. By contrast, involuntary medication of *competent* mentally disordered persons is regularly legally impermissible. A second class of cases concerns persons who present an imminent threat of some gravity to themselves or others. Irrespective of competence, such threats can be averted by the necessary means, including mind-interventions (e.g. sedation). They are justifiable, quite uncontroversially, by police powers or self-defence.

Our present interest, however, lies beyond these cases and concerns forcible mindinterventions, for the justification of which advocates invoke the advancement of *other* societal ends, particularly those related to punishment. The United States Supreme Court addressed this issue with respect to two case types: In *Washington v. Harper*, it upheld the involuntary long-term administration of psychotropic

medication to a *competent* but "gravely disabled" schizophrenic and dangerous inmate to maintain prison security, provided it is in his best medical interest[10]. As the threat was not imminent, the justification in *Harper* can be understood as a conjunction of the state's right to prevent harm to others under the special conditions of confinement, *and* paternalism. The *Harper* decision was controversial and not unanimous. The dissenting opinion criticised the majority for undervaluing the "defendant's liberty interest" to refuse mind-altering drugs, which merits the "highest order of protection",[11] but did not further elaborate on it. Instructive for the present inquiry is a brief discussion as to whether the state "might seek to compel Harper to submit to a mind-altering drug treatment program as punishment for the crime".[12] Dismissively, the Court wrote: "drugs may be administered for no purpose other than [medical] treatment".[13]

In *Sell v. United States*, the Supreme Court had to decide whether forcible administration of antipsychotic drugs to render a defendant competent to stand trial unconstitutionally deprived him of his liberty to reject medical treatment.[14] Unlike *Harper*, *Sell* was not presumed to be dangerous, and whereas he was not competent to stand trial, he was to refuse treatment. The court recognised a "severe intrusion" with a "significant liberty interest" to refuse mind-altering drugs, but did not clarify the scope or significance of this interest. It held that:

> "the Constitution permits the Government involuntarily to administer antipsychotic drugs to render a mentally ill defendant competent to stand trial on serious criminal charges if the treatment is medically appropriate [i.e., in the patient's best medical interest], is substantially unlikely to have side effects that may undermine the trial's fairness, and, taking account of less intrusive alternatives, is necessary significantly to further important governmental trial-related interests"[15]

A recent landmark decision of the German Constitutional Court points in the same direction.[16] A mentally disordered convict was detained after he had served his sentence as he was still presumed dangerous. The Court had to decide whether he could be coercively rehabilitated with psychotropics to render him fit for release. It held that coercive mind interventions can interfere with the core of the personality and are permissible if and only if the person is incompetent to make treatment decisions *and* the treatment is in his best medical interest. The ruling allows the inference that the mandatory treatment of competent persons for non-medical purposes violates constitutional rights.

In these decisions, the courts held that neuro-interventions interfere with a "significant liberty interest", without further defining its meaning or scope. Moreover, the courts held that this interest can be outweighed, in exceptional cases, by compelling governmental interests, a right to a fair trial and

[10] Washington v. Harper, 494 U.S. 210 (1990).
[11] ibid, 237.
[12] ibid, 241.
[13] ibid, 243.
[14] Sell v. United States, 539 U.S. 166 (2003).
[15] ibid, 167.
[16] Decisions of the German Constitutional Court, BVerfGE Volume 133, pp. 112-143.

prison security. Of present concern is whether moral rehabilitation can constitute a furthering of these interests. Apart from scholarly criticism of the decisions,[17] one significant aspect speaks against drawing analogies to this end. Although *Harper* and *Sell* were competent to withhold medical consent, both were afflicted with severe mental disorders. Even though societal grounds were invoked to justify interventions, they targeted mental disorders, were essentially therapeutic in nature and in the best medical interest of offenders. Thus, I suggest that these rulings are best understood as permitting states to pursue societal interests through interventions if (and only if) they are medically indicated. If interventions run against best medical interests, the law, as formulated in *Sell* and *Harper*, opposes them. This poses a problem for neurorehabilitation. As many methods, such as pharmaceuticals, bear medical risks without conferring health benefits, they are not in the best medical interest (in a purely medical cost-benefit framework, any non-therapeutically indicated but potentially harmful intervention runs against the best medical interest). Neurorehabilitation follows the calculus of trading potential setbacks to health for moral improvements. These interventions would be impermissible for contradicting the best medical interest of offenders.

Surely mental disorders and moral deficits overlap in a vast grey area, which may expand inasmuch as psychiatric categories – never free from moral or social considerations – expand as well. Onward from *Plato*, views of crime as expressions of a "diseased soul" have been fashionable.[18] Given that a high percentage of today's prison population suffers from mental disorders, one may argue that they should be treated forcibly if the psychological source of their transgression constitutes a disorder (e.g. antisocial personality disorder). However, this presumably overstretches the Supreme Court's reasoning and appears as a pretext: moral rehabilitation would come in the disguise of medical therapy, subtly drawing upon the latter's benevolent association with curing and caring, its paternalistic justification and the value of health. This line of argument is dishonest. Although politicians, lawmakers and corrective institutions have pursued it in the past and will likely do so in the future, one should insist that medical interests of an affected person are not necessarily identical with the moral interests of society. Crime and deviance are mainly interpersonal conflicts, not health problems. It might be conceptually possible to define "mental disorder" in a way that encompasses social deviance and moral problems without getting uninterestingly broad and capturing all sorts of negative setbacks in life. However, of interest here is not the conceptual but the normative. The questions is whether justifications for interventions against the will of affected persons which cure disorders extend to interventions that render them more morally favourable. A forceful reason against this is that the meaning of health seems more easily objectively definable than of moral rightness (at least, if one shares relativist premises). Also, imposing cures on refusers seems more persuasively

[17] E.G. Schultz, 'Sell-ing Your Soul to the Courts: Forced Medication to Achieve Trial Competency', Akron Law Review 38 (2005), 503; A.R. Dias 'Just Say Yes: Sell v. United States and Inadequate Limitations on the Forced Medication of Defendants', Southern California Law Review 55 (2003), 517.

[18] Plato, 'Gorgias', (Oxford: Oxford University Press, 1979), at 478a; 504; Adrian Raine, 'The Psychopathology of Crime', (San Diego: Academic Press, 1993).

justifiable than alterations of one's moral character. Thus, justifying mandatory moral improvement as the ailment of an illness conflates paternalistic health concerns with societal interests in moral conformity.[19] Speaking of "medical correctives", "treatment" or a "cure for crime" is, thus, unfortunate as it semantically blurs two domains – the medical and the moral – that should be kept separate for analytical purposes. In the following, to be clear, I am only interested in justifications of mandatory rehabilitative mind interventions not based on medical benefits.

In the search for legal precedents for this case-type, it pays to revisit debates of the 1970s,[20] prompted by then-emerging interventions such as behavioural modification, drugs, psychosurgery and, already back then, electric stimulation of the brain.[21] Not unlike contemporary advocates of moral enhancement, *Richard Delgado* called for the "psychocivilisation of men" through "pacification of the brain".[22] *B.F. Skinner* envisioned the "non-punitive society" that replaces aversive means with positive reinforcements.[23] *Anthony Burgess* wrote *A Clockwork Orange*, not only as a work of dystopian fiction but also to engage with the ambivalence of aversive conditioning programmes to which prisoners were subjected *realiter*.[24] To give you an impression: In California, always a forerunner in technology and mind-altering substances, convicted felons were administered anectine, a non-painful substance paralysing the body and suspending respiration for two minutes. Oxygen was simultaneously administered to avoid anoxia. While paralysed and frightened, offenders were counselled that their unacceptable behaviour must stop and that subsequent behaviour would be met with similar aversive treatment. This rehabilitative method applied the principles of *Pavlovian* conditioning, pairing criminal conduct with negative organic states (paralysis), so that the person, through subconscious learning mechanisms, seeks to avoid its future repetition.[25] In a different experimental treatment, films of shoplifting were shown to a habitual shoplifter. When items were stolen in the film, electroshocks were administered. Allegedly, it had the following outcome: "Treatment was successful. The patient finally stopped shoplifting, and she reported uneasy feelings of being watched whenever she entered a

[19] The relation between mental illness and morality is, of course, complex. The need to exclude mainly moral and social problems from the category of health is controversial but is shared by the main position in psychiatry, as expressed in the exclusionary conditions in the definition of mental illness, e.g.: "Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders", unless the deviance results from a dysfunction. American Psychiatric Association, DSM-V (2013), p. 20.

[20] Among the legal papers still worth reading today are M.H. Shapiro, 'Legislating the control of behavior control: Autonomy and the coercive use of organic therapies', *Southern Cal. Law Review* 47 (1973), 237-356; B.J. Winick, 'Legal Limitations on Correctional Therapy and Research', Minnesota Law Review 65 (1981), 331–422; D.B. Wexler, 'Of Rights and Reinforcers', San Diego Law Review 11(1974), 957-971.

[21] See e.g. E.S. Valenstein, 'Great and desperate cures', 1986; 'Brain Control', (New York: Wiley, 1973); J.A. Mills, 'Control: A history of behavioral psychology' (New York: NYU Press, 1998).

[22] J.M.R. Delgado, 'Physical Control of the Mind: Toward a Psychocivilized Society' (New York: Harper, 1969).

[23] B.F. Skinner, 'The non-punitive society', Japanese Journal of Behavior Analysis 5 (1990), 98-106; cf. Wheeler (ed.) 'Beyond the punitive society', (San Francisco: Freeman, 1973).

[24] B. Newman, 'A Clockwork Orange: Burgess and Behavioral Interventions', Behavioral and Social Issues 1991, 61-70.

[25] R.G. Spece, 'Conditioning and Other Technologies used to "treat", "rehabilitate", "demolish" prisoners and mental patients', Southern Cal. Law Review 4 (1972), 635 ff.

store".[26] In Connecticut, paedophiles were exposed to pictures of naked children and given shocks to the groin area when their bodies showed signs of arousal.[27]

These are illustrative examples of real experiments with coercive rehabilitative mind-interventions. Some programmes were halted due to public outcry; others were struck down by courts, not only because the interventions were experimental and painful but because they were considered "impermissible tinkering with mental processes" or "cruel and unusual punishment".[28] California amended its penal code. Section 2670, still in force, states that:

> "all persons, including all persons involuntarily confined, have a fundamental right against enforced interference with their thought processes, states of mind, and patterns of mentation through the use of organic therapies; … this fundamental right requires that no person with the capacity for informed consent who refuses organic therapy shall be compelled to undergo such therapy."

As this section seems to apply to contemporary successors of "organic therapy", Californian law explicitly outlaws mandatory moral neurorehabilitation.

These examples from US and German jurisprudence provide a rough picture: Societal ends might override the refusal of competent persons in exceptional cases and to further important state interests insofar as interventions promote the best medical interest. Outside of therapeutic contexts, by contrast, when reforming "criminal minds" stands in the foreground, the right to freedom of mentation and the protection of the brain as the locus of thought and personality seem to prevail.

One might object that the dark age of psychosurgery and aversive conditioning cannot provide precedents, as these methods are indeed cruel and ineffective, neither of which is true of modern neurorehabilitation. However, the normative argument of courts and scholars did not solely depend on the side effects of particular methods. It heavily drew upon the idea that the mind is a specially protected place. In fact, one can diagnose an interesting contrast between contemporary arguments and those from the 1970s. Back then, concerns revolved around ideas of "freedom of mentation", whereas today, health is seems to be the paramount value. That might be telling of our times.

II. Human Rights against Rehabilitation

Let us now turn to substantive arguments. Should states have a right to rehabilitate offenders against their will? In the following, I shall engage with the reasons for and against mandatory rehabilitation. I wish to highlight the set of rights of offenders who oppose mandatory neurorehabilitation. While it

[26] ibid, 655.
[27] Knecht v. Gillman, 488 F.2d 1136 (8th Cir. 1973).
[28] Mackey v. Procunier, 477 F. 2d 877, 878 (9th Cir. 1973). The decision that ended the psychosurgery of prisoners in the US, Kaimowitz v. Michigan (Jul 10th 1973 /73-18434), should be mentioned as the court addressed–and dismissed–the still relevant question whether prisoners can give valid consent to such treatments while imprisoned.

seems evident that it interferes with the rights of offenders, details about these rights, their strength and importance are not often found in the current debate. The US Supreme Court, as well as the German Constitutional Court, only speaks vaguely about a "significant liberty interest" against neuro-interventions. However, any discussion of the permissibility of rehabilitative interventions is strikingly incomplete – if not impossible – without a firmer understanding of the rights they affect. The right to health, which features prominently in today's judgments, is, I wish to suggest, not the main or most relevant right. Negative health effects are only contingent side-effects of neurorehabilitation. For the sake of argument, let us therefore assume that neurorehabilitation is free from negative health-related side effects. This opens the view to other effects in need of justification.

In the spirit of the arguments of the 1970s, it seems that the central interference pertains to what I shall outline in this section, a right to mental self-determination. This right stands behind the unspecified "liberty interest". Its neglect promotes the often encountered and uncritical attitude that rehabilitation, as such, does not face serious justificatory problems, which in turn leads to the revival of the crude *Hegelian* slogan of a "right to be punished". No such right exists. Properly understood, it means in the present context that persons have a *right against rehabilitation*. But there is no need to invent such a right. Some of the strongest, yet ill-defined, existing human rights protect persons against being subjected to rehabilitative mind-interventions. What are they? This, again, partially depends on contingent matters such as the nature and mode of operation of interventions. As we are not evaluating particular methods here, some broad observations must suffice. Effective neurorehabilitation alters moods, emotions, beliefs, thoughts and other cognitive or emotional processes of offenders. This raises the question of how the law protects these mental properties. And it leads to a general shortcoming of the law. The legal protection of the mind is a topic which has not yet been systematically addressed by legal scholarship or jurisprudence. It is unclear whether and which mental states or processes are protected against which kind of interference. However, there are traces of a protection of the mind in accepted human rights, which largely do not have a scope of application in jurisprudence. To contribute to the development of a mind-protecting right, I shall portray these rights potentially implicated by neurorehabilitation and provide a brief defence of their importance. I wish to emphasise that this exposition expresses my construal of these rights. They are not necessarily understood in this way in every jurisdiction, nor regularly applied to rehabilitative measures. So, the following interpretation seeks to contribute to human rights theory over mind-interventions by shedding some light on central rights currently in the dark in legal discourse.

1. Bodily Integrity

Apart from the negative side -effects of medication, the forcible administration of neurorehabilitation, such as psychoactive substances (e.g., injection), inflicts bodily harm. Although administration might not be very intensive (e.g. a pinprick), some psychiatric patients subjected to forcible medication

report it as traumatising.[29] Many persons are understandably fearful of being exposed to unfamiliar psychoactive substances, of "losing their mind" or control over themselves. This fear aggravates the invasiveness of an injection that may appear harmless on the surface. However, law enforcement regularly has to resort to physical violence. Inflicting bodily harm is not *per se* unjustifiable. But nonetheless: we would close our eyes to the real interference to which neurorehabilitation amounts if we remained on the evident, external and bodily level and were concerned about pinpricks only. Neurorehabilitation primarily alters minds.

2. Absolute Limits to Punishment

Constitutions and Conventions categorically ban specific forms of punishment. In the US, the 8th amendment prohibits "cruel and unusual punishment"; Art. 3 of the European Convention outlaws "inhuman or degrading treatment or punishment". Although these absolute limits have emerged as responses to particular practices of the past, their standards are evolving. Whether neuro–rehabilitations cross high thresholds remains speculative. Unless painful, they will likely not be considered torture. But some aspects warrant analogies. The wrongness of torture does not merely lie in its painfulness (inflicting bodily harm is sometimes permissible) but, rather, in breaking the victim's will from within:[30] Torture turns the internal mechanisms of a person against herself; it appropriates a person's susceptibility to pain and its overpowering psychological force to overcome mental resistance. It forces the person to dissociate herself from her body – or alternatively, from her will. In this sense, torture intervenes in central aspects of a person's self-relation. One could make the case that some neurorehabilitations work similarly: They overpower the cerebral configuration and mental self-control of a person and turn physiological mechanisms against herself. One's own body thereby becomes the source of the attack. This distorts or corrupts a person's self-relation.

Similar elements have been used by the ECtHR in defining inhuman and degrading treatment: "breaking physical or moral resistance" or "driving the victim to act against his will or conscience" are parts of the court's definitions.[31] Moreover, *brainwashing* is often named next to torture. An imprecise term itself, it does not necessarily involve unbearable pain or overwhelming force but can consist in severe psychological manipulation through methods such as inducing and exploiting fear, anxiety, the need for attachment, social isolation, or disorientation.[32] So, although most neuro–rehabilitations neither constitute torture *sensu stricto*, nor reach its level of severity, some forms might come close. This will depend on contingent properties of particular interventions and may require finer definitions of torture or related terms. But even if neurorehabiliations do not qualify – their proximity to absolute prohibitions is indicative of their severity, and the thresholds justifications have to pass.

[29] Jarrett et al. 'Coerced medication in psychiatric inpatient care: literature review', Journal of Advanced Nursing 64 (2008), 538-48.
[30] Insightful: David Sussman, 'What's Wrong with Torture?', Philosophy & Public Affairs, 33 (2005), 1–33.
[31] E.g. Keenan v. UK (April3rd, 2001–27229/95) at 110; Jalloh v. Germany (July 11th 2006–54810/00).
[32] K.E. Taylor, 'Brainwashing: The science of thought control', (Oxford: Oxford University Press, 2006).

3. Freedom of Thought and Conscience – (Cognitive liberty)

A couple of rights pertain to the mind more specifically. These rights, however, are poorly developed in legal scholarship and are neither properly defined nor applied in jurisprudence. These rights can be traced back to, and express, the liberal premises outlined above: the freedom of the human mind, even with respect to deviant or criminal thoughts. Freedom of thought and freedom of conscience are the most famous ones. They are core human rights, guaranteed by every treaty and covenant.[33] Together with the right to hold opinions without interference,[34] they encapsulate the notion that some parts of the inner world of the person have to remain outside governmental control. Although not precisely defined, the scope of the rights comprises the content of one's thoughts and the mental processes that form them, especially beliefs and opinions in moral, political or religious matters. For historical reasons, this protection is called freedom of conscience. "Conscience" denotes the psychological faculty to form and revise moral beliefs. The right emerged from times of incessant religious strife, in which every side claimed the divine authority to alter the religious beliefs of others. The acceptance of others' inner processes as inviolable formed the basis for political toleration. Being entitled to think for oneself and to form opinions independently is perhaps the central claim of the Enlightenment. Freedom of thought and conscience express deep scepticism over objectively correct religious or moral beliefs and the mandate of the state to enforce them. Whereas conscientious actions in the external world can be restricted, the inner side of the right, the so called *forum internum*, enjoys the strongest level of protection in human rights law. According to many, these rights are absolute, so that on principle, interferences cannot be justified.[35]

It then becomes crucial to define which interventions constitute interference. For lack of cases, there are only a few legal precedents. The scholarly literature views forceful or coercive alterations of opinions or their formation as illegitimate interferences but widely assumes that such interventions are factually impossible.[36] Accordingly, the right is almost without practical application. Interpretation and application of these rights require developing a framework to assess which interventions are severe enough to constitute interference. Given the strength of the right, not every governmental influence on the morality of citizens qualifies. States shape the morality of people in *prima facie* legitimate ways, such as moral education or public awareness campaigns, and the right cannot be implicated whenever state officials engage with citizens over moral matters. However, such influence

[33] Manfred Nowak, 'CCPR Commentary', (Kehl: Engel, 2005); L.M. Hammer, 'The international human right to freedom of conscience', (Aldershot: Ashgate, 2001); J.C. Bublitz, 'Cognitive Liberty or the International Human Right to Freedom of Thought', in: Clausen/Levy, Handbook of Neuroethics (Dordrecht: Springer 2014) 1309–1333.

[34] Art. 19 Universal Declaration of Human Rights; Art. 10(1) ECHR.

[35] In times of public outrage about people with deviant sexual fantasies, it should be underlined that freedom of thought also covers morally condemnable thoughts. It is not a criminal offence to have paedophilic thoughts (or dispositions). Prohibited–and grounds for public concern–are respective actions.

[36] Summarized in J.C. Bublitz, 'Freedom of thought in the age of neuroscience', Archiv fuer Rechts- & Sozialphilosophie 100 (2014), 1–23. .

must face limits.[37] Without drawing them more precisely here, invasive rehabilitative methods that tinker with neuronal processes underlying moral opinions, that do not rely on the persuasiveness of rational argument but, rather, re-programme citizens' brains against their will, impinge upon the right to free thought and to form and hold opinions without interference. In view of contemporary human rights theory, they are, therefore, prohibited across the board.

4. Right to Privacy or Personality

Furthermore, some jurisdictions explicitly protect the personality or character of a person; others capture personal and private matters under the rubric of privacy.[38] To different degrees and in various strengths, these rights comprise core aspects of the person as her character or identity (not in the numerical sense), her self-conception as well as her public image. Although concepts such as authenticity can be called into question on metaphysical grounds, this does not undermine the scope of the legal protection.[39] It is hardly deniable that every person possesses some characteristics central to who she is, in view of others and herself. Considering the central traits of a person as off-limits for state interventions does not imply dubious presumptions of an unalterable, innate, individual essence. The reason for this protection is the idea that every person has an original, i.e. pre-positive (and perhaps inviolable) right to herself. The way a person is, wants to be, how she perceives of and relates to herself, is an intimate and self-regarding matter. Its protection does not require approval of one's characteristics by the community at large. *Thomas Douglas* concedes that some moral characteristics can be considered "fundamental traits *par excellence*".[40] Immoral dispositions may, thus, fall under strong human rights protection. If interventions are capable of transforming central traits, their mandatory use has to be measured against the high standards of the protection of personality or privacy.

5. Respect for Autonomy, Agency, and the Person

Although not explicitly enumerated as a human right, respect for autonomy is a general principle of liberal legal orders. It is implied in the demand of respect for everyone as a free and equal citizen, a central idea of the social contract tradition and a premise of liberal orders. Autonomy is a multifaceted concept, often invoked without much precision. Three interrelated senses can be disambiguated here: For one, autonomy is the normative status of a free person in the sense that her decisions have to be respected and that she can be held responsible for her actions. Respecting this status entails not undermining the factual capacities necessary for autonomy. However, it is unlikely that

[37] With respect to drug rehabilitation, see R.G. Boire, 'Neurocops: The Politics of Prohibition and the Future of Enforcing Social Policy from Inside the Body', Journal of Law and Health 19 (2005), 216-257.
[38] Jill Marshall, 'Personal freedom through human rights law?' (Leiden: Nijhoff, 2008).
[39] J.C.Bublitz/R. Merkel, ´Authenticity and Autonomy of Enhanced Personality Traits´ 23:6 (2009), 360-74.
[40] Thomas Douglas, 'Moral Enhancement via direct emotion modulation: a reply to Harris', Bioethics 27 (2013), pp 160–168.

neurorehabilitation undermines such capacities.[41] Even severe interventions will strive to preserve autonomy in this sense; they do not aim to compromise rational and decision-making capacities.

Nonetheless, manipulative interventions can infringe upon autonomy in a *relational* sense in which it is contrasted with heteronomy. Vis-à-vis a manipulator who exerts heteronomous influence upon her, a person cannot be deemed autonomous. Here, autonomy means setting ends for *oneself*, to be independent of another's will, and in *Kantian* terms, to be a moral self-legislator. This sense of autonomy is undermined by asymmetrical power relations and interventions forcibly altering moral attitudes or behaviour.

Autonomy further relates to self-control. In the tradition of German Constitutional Law, one can speak of respect for a person as a subject. Debates over "agency" have a similar reference point. Respecting a person as a subject (her agency) implies respecting her as a self-controlling being. Mind-interventions may undermine self-control by accessing mental functions in ways which bypass control. Inner means to resist forcible neuro-interventions are weak and can likely be overpowered by stronger interventions. Neurorehabilitation may, thus, undermine self-control over targeted mental aspects, or mental autonomy.[42]

Accessing and altering the mind at the neurophysiological level marks a fundamental difference in the way to engage with another person:[43] One can engage at the level of communication, which is, in general, respectful of the other as a self-controlling being and which appeals to thoughts, opinions or emotions. By contrast, one can engage with the other at the level of the brain alone, which neglects contents of thoughts, emotions or personality traits, and only deals with neuro- and biochemical processes. Such interventions do not engage with the other as a self-controlling being but, rather, seek to overcome mental resistance or bypass mental self-control. Stimuli do not engage with moral flaws in the realm of reason but treat their causes at the physical level, reducing the content of the mental world of the recipient to brain chemistry. Note that there is no crude notion of mind-brain dualism at play here. The problem is not that interventions work at the physiological level – all interventions do. The problem is, rather, that they work *only* at this level and thereby circumvent the central element of a subject, or a person: her inner, subjective side, higher-level mental process and conscious content. Altering the mental machinery at the neurophysiological level alone is objectifying and disrespectful of the targeted person as a rational and self-controlling being and, thereby, of her as an autonomous subject.

---

[41] Of course, this depends on the conditions of autonomy: some historical theories consider persons as non-autonomous after intensive manipulations, cf. Vincent, 'Restoring'.

[42] For more on the notion of mental autonomy, cf. J.C. Bublitz, 'Moral Enhancement and Mental Freedom', Journal of Applied Philosophy 33(1) 2015, 88-106.

[43] This claim is strongly contested in neuroethics. See Neil Levy, 'Neuroethics: Challenges for the 21st century' (Cambridge: Cambridge University Press, 2007); Focquaert/Schermer, 'Moral Enhancement: Do Means Matter Morally?', Neuroethics (2015), 139–151; Bublitz/Merkel, 'Crimes Against Minds', Criminal Law & Philosophy 8 (2014), 51-77. Shaw 2014,

6. Mental Integrity

The aforementioned rights are not yet firmly defined in human rights law. Because their scope is unclear, they may not capture all possible interventions. Those interventions that fall short of interfering with them still impinge upon the weaker, and limitable, right to mental or psychological integrity.[44] Although recognised in human rights treaties, its scope and meaning have not been fully formulated. It is the minimal right against which any mind-intervention has to be tested.

7. Preliminary Conclusion

In conjunction, the rights to freedom of thought and conscience, to privacy and personality along with the respect for autonomy, provide protection against neuro-interventions. Some of these rights prohibit targeting the opinions, moral beliefs or central characteristics of the person. Others primarily oppose manipulative, control-bypassing, or objectifying means of treating other persons. Together, these rights protect persons against being stripped of constitute elements of themselves: autonomy, personality, thoughts, moral and religious opinions, in ways that circumvent or undermine mental self-control. The scope of the protection partly coincides with what many scholars consider human dignity. It surely has a *Kantian* tone and captures, I suppose, what authors such as *Herbert Morris* mean by the demand to "treat individuals as persons".[45]

Without getting entangled in controversies over dignity, I shall refer to this set of not well-defined rights that protect the mind summarily as the right to mental self-determination. Any rehabilitative method that interferes with it has to be justified. Whether justifications succeed depends on the intrusiveness of the intervention and the strength of opposing rights. According to contemporary human rights theory, interventions altering beliefs, opinions or thoughts cannot be justified at all; the protection is absolute. Consequently, most neurorehabilitations would be impermissible *tout court.* While I have sympathies for a strong human rights protection of the human mind, I do not wish to commit myself to the claim of absolute protection of the human mind. To be persuasive, it must be elaborated upon and specified quite a bit. After all, we change each other's minds all the time. Nonetheless, I espouse the idea that the protection of the mind must be strong and that justifications for interferences face peculiar problems as they contradict liberal premises which raise the bar considerably. I will return to this point at the end of the chapter.

With respect to assessing the intrusiveness of particular methods, it is helpful to rank interventions on a spectrum in light of contingent empirical matters, depending, for instance, on how strongly they interfere with free thought, the degree to which they undermine self-control, or the intensity and

[44] Protected in the jurisprudence of the ECtHR under Art. 8 (privacy) and explicitly enumerated in Art. 3(1) of the European Charter of Fundamental Freedoms.
[45] Herbert Morris, 'Persons and Punishment', in: J. Murphy (ed.), Punishment and Rehabilitation (Belmont: Wadsworth, 1995), 74-93.

duration of effects. This ranking should include indirect methods (e.g. psychotherapy). On one side of the spectrum lie ordinary and everyday ways to influence others to become morally better persons, which may not even qualify as interference. Direct brain interventions bypassing mental control fall at the other end. Of course, the chapter alluding to *Foucault* in the title has to note that social power is ubiquitous. Distinctions between means and methods do not purport that only some particularly powerful interventions are capable of changing minds. No, even ordinary social structures are. But the normative challenge is to delineate permissible from impermissible influences. For this, the mode of operation and effects of an intervention are key criteria. Interventions providing new mental skills (cognitive-behavioural techniques) or enhanced self-control through reducing emotional urges may fare better than those altering opinions or beliefs. Recent debates about moral enhancement have focused on emotional manipulations. Although emotional dispositions might not enjoy the same level of protection as thoughts or conscience, emotional alterations could change thoughts and conscience indirectly. A pill strengthening feelings of guilt, for instance, likely also impacts opinion formation.

Jurisdictions may assess the strength of protection afforded by the right to mental self-determination and the intrusiveness of particular interventions differently. Some will consider neuro-interventions to violate mental self-determination across the board; others may be tempted to allow mandatory neurorehabilitation in exceptional circumstances, especially when prompted by public outrage over supposed softness with offenders. These latter jurisdictions, however, have to acknowledge that affected rights rank among the strongest guarantees in international human rights law and probably in domestic constitutional law. I will provide some arguments in support of their importance in the final section. Preliminarily, I shall conclude that the majority of mandatory neurorehabilitation interferes with some of the strongest human rights. The threshold for justifications is, thus, high and considerably higher, I would like to suggest, than the right to health.

III. The State's Right to Rehabilitate?

Let us now turn to the other side, to arguments in favour of mandatory neurorehabilitation. I seek to show that the impression that rehabilitation does not face serious justificatory problems is misleading. The case for mandatory neurorehabilitation is much weaker than often suggested. Commonly, justifications for mandatory rehabilitations approvingly note that rehabilitation is a widely accepted penological goal and then embark on a cost-benefit analysis in which it fares considerably better than hard treatment (e.g. incarceration). But this line of reasoning is insufficient and incomplete because it falls short of explaining from where the state's power to rehabilitate offenders against their will emerges.

1. Aims, Interests, & Rights

More concretely, arguments in favour of mandatory neurorehabilitation usually resort to variations of the following claims: (i) Moral rehabilitation reduces recidivism. (ii) Rehabilitation is among the accepted penological aims (perhaps the least controversial of them). (iii) Pursuing penological aims belongs to the legitimate interests of the state. (iv) Furthering legitimate aims of the state may justify interferences with rights of citizens, even hard treatment such as long-term imprisonment. (v) If hard treatment as imprisonment is justifiable, so are less restrictive and potentially more effective alternative measures. (vi) Rehabilitation is less severe, more humane and potentially more effective than imprisonment (the "lesser of two evils"). (vii) If imprisonment is justifiable, so is rehabilitation.

The argument may appear familiar and straightforward, but it passes over three problematic aspects too quickly: the use of "moral" in (i), the idea that interests of the state *can* justify interferences with rights (iii and iv), and the comparative "lesser of two evils" argument (v to vii). More precisely, what does it mean to say and what are the implications of saying, that rehabilitation is a penological aim and furthers legitimate state interests? It is helpful to distinguish a weak and a strong reading.

The weak version states that morally rehabilitating offenders is, by itself, a legitimate end for the use of state powers, i.e., it falls within the state's competence, it is something governments can rightfully pursue. How could one find any objection with that? Well, powers of the state are limited, and moral improvement may fall outside of its purview. One reason is the conflation of morality and law. Whether both are strictly separable belongs to the venerable controversies of legal philosophy (positivism v. naturalism), and which moral views should enjoy legal protection concerns the neutrality of the state. But irrespective of both issues, the coercive apparatus of the state should only be used to enforce those moral values which are embedded in the law, e.g., as rights of individuals or the community, or expressed in constitutions. Accordingly, the term *moral* rehabilitation is too broad, (i) has to be understood more narrowly. If at all, states can only reform offenders into law-abiding, not morally improved citizens. Accordingly, only *criminogenic* psychological factors, not moral deficiencies, are permissible targets of rehabilitation.[46]

[46] In this vein, the German Constitutional Court ruled in the 1960s that making citizens live a morally "better", happier or more productive life is not among the tasks of the state, unless they present threats to others or themselves. Moral enhancement cannot, therefore, justify interferences with rights of competent and law-abiding, yet non-virtuous, morally flawed persons (Decisions of the German Constitutional Court, BVerfGE Volume 22, p. 181). This line of thinking corresponds to the perspective of classic liberal conceptions of law such as *Immanuel Kant's*. The purpose of the state is to regulate external conduct, not internal (mental) states. The legitimacy of the state–and its coercive powers–derives from the necessity to regulate conflicts between persons, and such conflicts can only arise in the external world where the actions of one can collide with those of others. In modern words, the proper domain of coercive state regulation lies in the solution of inevitable conflicts over external resources. Mental states or the character of a person, by contrast, cannot conflict with freedoms of others, so restrictions are unnecessary and illicit. In this concept of the state, its powers are restricted to securing reciprocal external freedoms. Morality is a private matter. Classic liberal models have lost much of their appeal. Republican or welfarist models–and the political conception of existing states–allow for the advancement of substantial goals, including public morality. One may grant states, as Joseph Tussman argued, a moral "teaching power", allowing them to arrange internal affairs to promote morality and rehabilitation ('Government and the mind'; New York: Oxford University Press, 1977). They may instruct

Moreover, invoking legitimate aims of the state as a justification for interferences with rights requires a stronger reading of what "legitimate aims" implies. According to it, aims are *pro tanto* justifications for interferences with rights. Only in this understanding can the state's interests in rehabilitation outweigh the rights of affected persons. This stronger sense is the one in which laypeople and lawyers often speak (and unfortunately think) about the relation between interests and rights. The problem with this understanding is that a *factual* interest is thereby elevated into a *right*. But rights and interests are different entities and cannot easily be pitched against each other, at least not without invoking further theoretical premises. Different normative frameworks operate in the background: Roughly, rights come from deontic, interests from consequentialist models.[47] Rights can come into conflict with other rights; and so can interests with other interests. But rights and interests are, at least *prima facie*, incommensurable. Therefore: factual interests of the state or benefits for the common good, large as they may be, cannot simply be taken as justifications for rights-infringements. This affects claims (iii) and (iv).

Ordinary legalese sometimes confounds rights and interests. To many, it does not sound wrong to say that citizens' rights can be overridden if governments show a "compelling interest". And regularly, speaking of rights and interests interchangeably is, indeed, harmless because a respective right underpinning the interest is easily construable. However, in particular cases, such as the present one, it is not. As this is a crucial yet often overseen aspect, let me illustrate the relation between interests and rights with an example: Suppose A is strongly interested in buying the house of his neighbour B. This is a legitimate interest of A, in the sense of being a permissible goal for him to pursue. Nonetheless, it does not entitle A to interfere with B's property right, e.g. through actions that lower the price of the house or render B more inclined to sell. Notably, A's factual interest in acquiring the house, and the good outcomes it may generate, are *not* balanced against B's right to keep his property. Even if A's interests were far stronger than B's, no entitlement to interfere with B's property right would arise. So, even legitimate and weighty interests of one party do not justify interferences with rights of others. B's legal position, the right to property, is not conditional upon the absence of stronger interests of others. B can freely decide to sell. In this sense, rights trump interests.

The same would be true if A were the state. Legitimate interests of the state do not, by themselves, undermine citizens' (and B's) rights. With respect to property in land, legal systems have expropriation clauses which allow appropriating land against the will of proprietors (often in exchange for a market price). If the conditions of expropriation obtain, the state can force B to sell. However, to

---

employees to be moral leaders, praise the morally virtuous, incentivise moral behaviour, design programmes and institutions under the guiding idea of moral improvement, provide moral education or engage the public with moral questions. But, importantly, granting such a "teaching power" does not ipso facto justify interferences with opposing rights of citizens. While the state may speak out on moral matters, citizens do not need to listen, let alone espouse what is said.

[47] This also relates to the controversy in legal theory over interest vs. will theories of rights. In law, a right denotes a position stronger than an interest.

invoke such clauses, the state needs to show that the acquisition of property is necessary to fulfil *other* duties, e.g. building critical infrastructure or protecting rare wildlife. The state's rights and duties to govern *in other domains* may necessitate and justify expropriation.

If we transfer this to neurorehabilitation, it becomes evident that the strong interests of the state in reforming offenders are, by themselves, insufficient to override their rights to mental self-determination. The state needs to show more, a right allowing and requiring interventions into the minds of offenders, which, importantly, cannot lie in a "right to rehabilitate". This would simply beg the question, because it already assumes that the state has a right over the minds of offenders. But this right, as part of mental self-determination, belongs to affected persons, not the state. And it entails a claim against the state to non-interference with minds. The state cannot, thus, simply claim to possess the same right over the same object. Claiming this would already run counter to the right of the person (as it would violate the non-interference claim). Thus, the right to mental self-determination blocks, if you will, the emergence of a similar right of the state over the very same object. Therefore, analogous to expropriation clauses, the state needs to show a *further, additional aim* within its competence that outweighs affected persons' rights to mental self-determination. Arguments merely referring to rehabilitation as a penological aim, or to the interests of the state, do not explain where this additional element comes from (iii and iv). By default, states do not have rights over minds of citizens. Recall the cases of *Harper* and *Sell*: In both, the state invoked additional interests in the rank of rights – to maintain prison security and the right to a trial – which required offenders to be in a sound state of mind.

2. Forward- and Backward-looking Aims of Penal Law

Arguments for mandatory rehabilitation thus have to demonstrate an additional legally protected interest of the state in the moral character of a person, *over and above the mere benefits of its improvement*. Many candidates are conceivable. Here, I can only comment on the most common ones; and it seems helpful to sort them according to their relation to forward- or backward-looking objectives of penal law. With a view to the future, states are obliged (and have a right) to protect citizens and to prevent future offences, traditionally the domain of police law. With a view to the past, states need to respond to offences, the traditional domain of criminal law and punishment *strictu sensu*. Let us assume *arguendo* that states possess a right to punish. In legal practice, the boundaries of forward- and backward-looking domains have become blurred. Penal sanctions are often understood (and justified) as prevention, whereas police powers are invoked for cases that call for criminal indictment (Guantanamo Bay is the symbol of the convergence between criminal law and preventive systems – long-term detention under punitive conditions without criminal trial). However, both

domains can and should be separated for analytical purposes as they are justified by different rationales and governed by different doctrines and principles.[48]

2.1. Forward-looking: Preventing offences

Can forward-looking aims provide us with the additional interest that may justify mandatory neurorehabilitation? Rehabilitative measures serve the future-directed objective to reduce recidivism. Its legitimacy derives from the protection of the public, which is among the *raisons d'être* of any state. States clearly possess a right to prevent offences. However, a different basic legal principle becomes relevant here: *the principle of the least restrictive alternative*. Because liberal states are bound to preserve civil liberties to the fullest extent possible, they have to interfere with rights of citizens in the least invasive way only. All else is excessive use of state powers. Future offences can be prevented through incapacitation and incarceration instead of mandatory rehabilitation. But is locking a person away less restrictive than mandatory moral rehabilitation, i.e. changing relevant mental properties against their will? The "lesser of two evils" and "treatment instead of punishment" claims (v and vi) draw their appeal from the former's appearance as *less* cruel and invasive. Whether this is the case partially depends on contingent empirical matters, which cannot be discussed here. More abstractly, it is hard to assess whether deprivations of liberty or alterations of one's moral character are less severe. As both affect different rights and interests, they seem hardly commensurable by objective standards. In terms of abstract strength, the right to mental self-determination outweighs rights against imprisonment. Accordingly, incapacitation through imprisonment should be conceived as the less restrictive option by default. But even without this claim, as a general rule, I suggest that in cases in which it is impossible to objectively identify the more burdensome measure, the rightholder has to choose which rights are more valuable to her, whether she prefers sacrificing physical or mental liberty.[49] The principle of the least restrictive measure thus demands that offenders be offered a choice between the two possible evils of undergoing character transformation or incapacitation. Once offenders can choose – and reject – neurorehabilitations, they cease to be mandatory, even if chosen only because the alternative appears worse. Thus, arguments in favour of coercive rehabilitation to prevent recidivism faces the objection that less restrictive alternatives are available: incapacitation, accompanied with the offer of neurorehabilitation.

The same is true for general deterrence. If punishment is justified because it deters other prospective offenders, the question is whether mandatory rehabilitation factually deters others, and whether less invasive alternative means to this end are available. The diminished amount of suffering caused by

[48] This is not to say that they might not be connected at a deeper level. The justification for the entire penal system may ultimately lie in future consequences (deterrence). Nonetheless, at the level of penological aims, forward- and backward objectives can be separated.

[49] There is a parallel discussion in the ethics of psychiatry over involuntary medication or seclusion. Patients report varying preferences; see Georgieva et al., 'Patients' Preference and Experiences of Forced Medication and Seclusion', Psychiatric Quarterly 83 (2012), 1.

neurorehabilitation to which advocates appeal is a justificatory weakness in terms of deterrence. Supposedly, incarceration and suffering has stronger deterrent effects. Only if it could be shown that neurorehabilitation is a powerful deterrent (more effective than other means) might it be justifiable under the idea of general deterrence. As this seems unlikely, the state can regularly pursue future-looking objectives with less invasive means and thus cannot justify mandatory neurorehabilitation.

2.2. Backward-looking: Retribution

Mandatory neurorehabilitation appears as a future oriented policy *par excellence*. Any argument in its favour seems to be connected to the beneficial future effects of moral reform (e.g. reducing recidivism). But as such preventive aims cannot justify coercive mind-interventions, justifications based on backward-looking considerations merit attention. Sometimes, neurorehabilitation seems to be understood *as a form of punishment*. Retributive punishment in a strict sense is concerned with the past, imposed in virtue of an offence (rather than some future state of affairs). Its objective, depending on one's favourite theory, is to condemn, redress or annul the offence, and to vindicate the right. Rehabilitation is often conceived as a means to alleviate the hardship and negative effects of retributive incarceration and to facilitate reintegration upon release. Therefore, rehabilitation is invoked as the main aim of *penitentiary system* or *correctional facilities*.[50] As a consequence, rehabilitation and retribution regularly appear together in practice. But this does not turn rehabilitation into a genuinely backward-looking penological aim. It is auxiliary or complementary to hard treatment, aiming to counteract its negative effects. But it cannot shoulder the justificatory burden (of punishing and imprisoning) on its own. Against this backdrop, it becomes evident why calls to simply replace imprisonment with rehabilitation are misguided. It would require changing normative foundations, but the justificatory grounds for hard treatment – retribution – *prima facie* do not apply to rehabilitation. A retributive argument for mandatory rehabilitation would have to argue that the offender, because of his offence, either *deserves* to have his moral character altered, or that a different retributive justification of punishment comprises rehabilitative sanctions. Can such an argument be made?

Jesper Ryberg suggests that retribution does not imply principled objections against replacing hard treatment with rehabilitation.[51] The family of retributive theories is large and diverse, perhaps even without a common core.[52] Unsurprisingly, then, some accounts may not *expressis verbis* object to replacing hard treatment with rehabilitation. It seems, however, that classic retributive theories imply a *specific relation* between the present and the past, between offence and retributive response. Thus, even if one grants *arguendo* the key controversial retributivist claim that offenders deserve something

[50] E.g. Art. 10(3) Covenant on Civil and Political Rights; Council of Europe, 'Recommendation on European Prison Rules' (2006), at 6; § 1.02(b) Model Penal Code; § 2 German Prison Act (StVollzG).
[51] Jesper Ryberg, in this volume.
[52] John Cottingham, 'Varieties of Retribution´, The Philosophical Quarterly 29, 116, 1979, pp. 238- 246.

in virtue of their offence, it has to be demonstrated that what they deserve is being forcibly subjected to neurorehabilitative programmes. In more general terms, the proposition would have to state: Because offender O has committed an offence, i.e. violated the rights of victims protected by penal law, O deserves alteration of his moral character.

I must admit that I fail to find the standard by which to evaluate this claim because the two elements lack any internal connection. What the offender did to the victims and what he allegedly deserves in return are entirely different matters, relating to different rights and interests. Retribution, however, draws on an equivalence of (or at least presupposes a strong link between) offence and response. Details depend on the retributive theory one favours. Without getting into specific accounts, one can observe that in retributive thought, punishment is primarily connected to the crime, not the criminal[53]. The aim of the retributive response is to assert the right or to nullify or negate the crime and its effects (rather than bringing about a state of affairs related to the criminal as a person). Advocates of retributive rehabilitation would have to explain how goals such as repayment or annulment were achievable through modifying the character of offenders. A standard retributive view would assert, to the contrary, that equilibrium is restored through the suffering of the offender. This essential connection between crime and sanction is most evident in the *urfom* of retributivism, *lex talionis*, which entails the perfectly understandable (though not endorsable) claim of equality between offence and punishment. What offenders did is what they shall get in response. During the evolution of penal theory and practice, especially through the influence of human rights, this equivalence has been modified, but not fully discarded. The reason why the torturer is not tortured anymore is not that equivalence was renounced but that particularly inhumane practices were. More generally, instead of responding to crimes with the same type of conduct, more humane surrogates in the form of fines and imprisonment were introduced and became the "currency" in which sanctions are made commensurable with offences. Nonetheless, with these modifications, the connection between crime and punitive response still persists and becomes manifest in principles such as "the punishment must fit the crime" or proportionality, cornerstones of just sentencing. The severity of the offence – in abstract type and concrete effects – sets the tariff and upper limits of what can be meted out to offenders. Anything over and above is considered excessive. Whoever accepts proportionality must accept an inner relation between offence and punishment. And this relation extends to the *kind* or *form* of sanctions. States cannot impose anything on offenders on the occasion of them having offended. Sanctions have to be non-arbitrarily related to what offenders inflicted on victims. Whereas one may argue whether fines and imprisonment observe this relation in every instance, neurorehabilitation falls squarely outside it.[54] The contention that the perpetrator deserves to have his moral character altered

[53] For the difference between criminal law systems based on offenders or on offences, see Georg Fletcher, The Grammar of Criminal Law, Volume One: Foundations (New York: Oxford University Press, 2007), p. 27.

[54] However, forms of neuro-interventions observing this relation are conceivable; e.g., interventions inducing suffering or remorse (cf. Purgh/Maslen, 'Drugs That Make You Feel Bad'? Remorse-Based Mitigation and Neurointerventions', Criminal Law & Philosophy 2015, doi: 10.1007/s11572-015-9383-0).

because of his wrongdoing hangs in the air. A deficient moral character might be among the *causes* of an offence. But what, in terms of desert, is the relation between causes and effects?

Alternatively, one might suggest making rehabilitation pain- or harmful. If anything, retributivism justifies inflicting suffering. However, that suggestion is misleading because the infliction of pain and the rehabilitative transformations can be analytically separated. Retributivism may justify the former but we are interested in the latter - in specific justifications for changing the moral character of offenders and the interference with the right to mental self-determination. The retributive justification for harm does not extend to rehabilitative effects.[55]

After all, a fully spelled out retributive theory likely stipulates some connection between offence and punishment. If so, the moral transformation of offenders fails to observe it. Under retributive premises, the state's right to punish, then, does not entail a right to subject offenders to neurorehabilitation.

2.3 Moral Reform as an aim *sui generis*?

Thus, the right of the state to rehabilitate offenders against their will arises neither from preventive nor retributive considerations. Yet again, one might ask why moral rehabilitation may not constitute an aim of its own, a justification *sui generis*. Why should society not respond to offences by morally rehabilitating offenders? As argued, although a legitimate aim of the state, rehabilitation does not self-evidently warrant interferences with rights of affected persons. Aims, by themselves, are no justifications for interferences. Calling for mandatory rehabilitation requires the stronger claim that the state has a genuine, legally protected interest in the morality of the person, over and above the prevention of future offences and retribution. More precisely, advocates have to present an argument that fulfils the following five conditions: It has to (1) establish a relation between rehabilitation and the past offence, without drawing on (2) retributive ideas such as desert, or on (3) the prevention of future offences, or on (4) the alleviation of medical conditions. Finally, the right of the state has to (5) outweigh the right to mental self-determination of offenders.

2.4. Paternalism

A remaining option might be the paternalistic route, according to which punishment is justified by the rights or interests of offenders themselves (rather than of the state or society). Technically, such an argument holds that offenders' rights against rehabilitation are outweighed by the rights and interests of the same person. I cannot engage with the peculiar and deeply problematic structure of paternalistic

[55] One might also invoke forfeiture theories. But even if offenders forfeit rights, it needs to be shown that they forfeit the specific right to mental self-determination. It seems more plausible that offenders, if anything, forfeit those rights which they have violated through their offence (of victims). If forfeiture is meant to carry more argumentative weight than merely justifying that states can punish offenders in any way they like, a specific relation between the offence and the forfeited right has to be established. So, again, the link between justification of punishment and rehabilitation is missing.

justifications here, but only wish to note that portrayals of punishment as in the best interest of the punished betray a deep sense of cynicism and tend to reverse the interests at stake. At least for liberal states, it is impermissible to replace the factual interests a competent person holds with different (and supposedly superior) interests of the community and declare them as the "real", "true" or "best-understood" interests of the person. A liberal state neither overrides subjective standards with supposedly objective ones, nor disposes of rights of persons against their will (unless the person is not competent to rationally define her interests and exercise her rights). This is not to say that moral rehabilitation cannot be beneficial in some sense – I suppose it often is – but not according to offender's standards. And they matter in justifications drawing upon the individual.[56] The community, not the offender, seeks mandatory rehabilitation, and we should not conflate their respective interests. With the paternalistic route forestalled, the remaining interests potentially justifying mandatory rehabilitation are those of the state.

2.5. Communicative Theories: *Duff*

A final justification might be found in communicative theories. I shall briefly (and in a coarse-grained way) address *Anthony Duff's* influential version.[57] As an ideal theory, it does not seek to justify actual penal practices but provides a standard against which they can be measured. The gist of communicative theories lies in the contention that offences provoke an expression of condemnation by the community. Punishment, including imprisonment, is conceived and justified as part of a "communicative process of argument and persuasion" between state and offenders.[58] This process seeks to instigate a process of self-reform, to "induce an appropriate change in the offender's attitudes" which involves repentance, feelings of guilt and remorse. Within the family of communicate theories, *Duff's* account allows for the most extensive influence on the moral character of offenders.[59] He proposes that the coercive apparatus should "try to reach offender's moral conscience" and character.[60] Offenders should not be "free to close their minds". The communicative aspect of imprisonment is that it provides an environment conducive to self-reflection and remorse; it renders

[56] Offenders' standards are not necessarily misguided: behavioural economics shows that offending can maximise interests of a fully rational *homo oeconomicus*, if long-term gains surpass costs (determined among others by the probability of being caught and the severity of the sanctions imposed).
[57] Anthony Duff, 'Punishment, Communication, and Community', (Oxford: Oxford University Press, 2001).
[58] Anthony Duff, 'Expression, Penance and Reform', in J. Murphy (ed.), Punishment and Rehabilitation (), p. 196.
[59] Cf. A.v. Hirsch in: 'Censure and Sanctions', (Oxford: Oxford University Press, 1993), ch. 8; and the debate between both, Anthony Duff, 'Punishment, Communication and Community' in: Matravers/Pike (eds.), 'Debates in Contemporary Political Philosophy', (London: Routledge, 2003), 387-407; and the following chapter by A.v. Hirsch, 'Punishment, Penance and the State', pp. 408-422; as well as Duff's reply, 'A response to von Hirsch', pp. 423-427.
[60] Duff, 'Punishment', pp. 396-97.

"communication more effective, forcing the criminals attention" onto his crime and may arouse moral emotions.[61]

If communication justifies incarceration, so one might argue, it may also justify mind-interventions supportive of the communicative process (we may call this "enhanced communication"). But notably, even *Duff* explicitly rejects the manipulation of offenders through drugs or psycho-techniques. Even if less painful and more effective than incarceration, they fail to

> "address the criminal as punishment must address her, as a responsible moral agent who can and should come to understand the moral implications of what she has done; they are not, as punishment must be, attempts to solicit and arouse her repentant understanding of her crime". Offenders must be "free to remain unpersuaded and unrepentant".[62]

Duff's rejection of neurorehabilitation is based on the image of penance in which true reform is *self*-reform, induced through painful moral emotions. Punishment has to appeal to understanding and address offenders as rational agents.[63] Coercively imposed transformations are unsuitable as they do not originate within the self. However, this distinction between internal and external sources, genuine forms of (self-) transformation and other ways, lacks persuasiveness. Why could there not be morality pills that create suitable *internal* conditions for communication, that "solicit and arouse understanding", or support empathy and repentance? Why should there not be a remorse pill?[64] Or pills that enable empathy or simulating different perspectives, which facilitate understanding? It seems that neurointerventions do not *necessarily* eradicate the freedom to remain unrepentant and may leave sufficient room for self-reform. Thus, I suggest that an argument in favour of mandatory neurorehabilitation as tools for enhanced communication and improved self-understanding might be made, *pace Duff*, on the basis of his model.

However, such a proposal falls prey to more general objections against *Duff's* account. To him, imprisonment is justifiable as a necessary and desirable form of communication. But to consider imprisonment merely as an annexe to – or suitable background conditions for – communication and self-growth misses essential parts of its nature and the rights it infringes upon. One may – I do – agree that offences provoke a communicative response and that offenders have a duty to participate, to receive and listen to censure and condemnation. Yet, it is doubtful that the kind of communicative process which offenders owe includes severe interferences with rights and psychological manipulation. Communication through prison walls amounts to more than a "dialogue" or a "process of argument and persuasion". It involves weakened mental resilience, increased susceptibility, deprivation, social isolation and feelings of solitude and a fair bit of desperation. These emotional and subconscious processes are, in a sense, *externally* induced by the prison environment. The

[61] Duff, 'Expression', pp. 194 ff.
[62] ibid, 195/197.
[63] Duff, 'Punishment', esp. sections 7/8.
[64] Purgh/Maslen, ´Drugs´.

psychological effects of total institutions raise the chances of penance and are harnessed precisely for this reason.[65] The malleability of the human mind through deprivation and counselling was proven, in extreme form, by what has been described as "coercive persuasion" in Maoist thought-reform camps.[66] Of course, *Duff* does not defend those. But what he considers communication is a far way from regular, let alone ideal (in a *Habermasian* sense) speech conditions characterised by argument and persuasion. Psychological manipulation, through prisons or pills, exceeds what is justifiable as a necessary communicative response to the offence. Such interferences are in need of further justification beyond communication. In other words: It is possible to communicate censure to offenders in ways respectful of their right to mental self-determination. A state's right to communicate, therefore, does not necessarily comprise permission to intervene into offenders' minds.

However, I would like to indicate that what I take to be the most promising possible field of application for coercive mind-interventions, apart from therapy, might lie in enabling offenders to better understand the perspective of victims and to take a self-critical view on oneself. Temporary administration of neurotools, perhaps as part of an augmented psychotherapy, might be conducive to this end. The normative argument, however, cannot be confined to communication. It would have to show that self-reflection and perspective-taking can be duties of citizens, at least in conflict situations, perhaps as part of a yet to be defined "fitness to be punished", and that these interventions are less worrisome than moral alterations in light of mental self-determination. But I must leave exploring this idea for another time.

IV. Conclusive remarks

In conclusion: The state has to demonstrate a right to rehabilitate, which has to be grounded in legally protected interests in the moral character of the person, over and above beneficial effects of its improvement. Whereas states surely have a right to prevent future offences, forward-looking objectives do not confer a right to mandatory rehabilitation, as offenders rights must be curbed in accordance with the principle of the least restrictive measure. In contrast to the supposition that incarceration is brutal whereas rehabilitation is humane, the former should be the default option, as incapacitation only infringes upon external liberties. Furthermore, backward-looking considerations do not provide a right of the state either. Retributivism implies a connection between offence and response, which neurorehabilitation fails to observe. While some forms of neurorehabilitation might be understood as enhanced forms of communication, it remains unclear why offenders owe participation in *such* communication. On balance, the case for the state's right to rehabilitate offenders is much weaker than it may appear at first glance.

[65] Foucault, 'Discipline and Punish'.
[66] Cf. Lifton, ´Thought reform´.

Surely, further interests of the state are conceivable, perhaps alongside alternative penal sanctions. Here, some room for creative arguments and considerations remains. For the time being, I can come up with only two further interests: financial resources and effectiveness. Neurorehabilitation might prevent crimes more effectively than incarceration, for instance, if offenders pose a threat within corrective institutions or if preventive detention is unjustifiable because potential crimes are only minor offences. Secondly, incarceration might be more costly than neuro-interventions.

Let us begin with the latter: Incapacitation is achievable through detention, which might be more expensive than neurorehabilitation. If such arguments from costs were successful, some of the strongest human rights would be restricted for financial reasons. Putting a price tag on such central guarantees runs counter to the idea of human rights. Regularly, financial reasons cannot be considered a compelling state interest to justify interferences with human rights, at least when negative rights are at stake (positive obligations – welfare entitlements – are surely a different matter). In other words: there is no right of the state to the less costly alternative. If states are unable or unwilling to pay for incarceration, they do not have to. But this results in the release of offenders, not permission to further interfere with their rights.

Secondly, eliminating dangerous behavioural dispositions may prevent crime more effectively than containing them. However, as *Harper* shows, threats within penitentiaries can be met with police powers, and that may include mind-interventions. More generally: the right to self-defence regularly outweighs the mental self-determination of attackers. However, averting imminent attacks – or maintaining prison security – is not in question here. Such interventions do not require character rehabilitation, but rather sedation and incapacitation. Their permissibility is neither new, nor intrinsically related to neurorehabilitation, nor rooted in the punitive powers of the state.

Moreover, some offenders may pose risks that are too low and vague to warrant preventive detention, yet we can surely predict that, statistically, some of these will realise. However, their realisation seems to be, primarily, a failure of other preventive measures. That some offences cannot be averted through reasonable means is a manifestation of the risks inherent in liberal societies, not a compelling argument to sacrifice their basic structure. If risks are too vague to warrant detention, they are equally so to justify interferences with strong human rights.

After all, then, it seems that the punitive powers of the state cannot justify mandatory neurorehabilitation. The case for such rehabilitation programmes is weaker and the rights that would have to be outweighed are stronger than often suggested. Of course, in the realities of politics and policies, there is a high chance that lawmakers are willing to impose mandatory neurorehabilitation programmes on offenders, in spite of the foregoing considerations. The public is, by and large, unwilling to spend resources on offenders and quick to devaluate their basic rights. Ultimately, the permissibility of mandatory rehabilitation programmes will depend – apart from contingent empirical matters – on the importance that courts assign to the right to mental self-determination. I would thus

like to conclude by sketching two arguments in support of its significance in liberal legal orders. It may not render the right absolute, but strong enough to prevail over most rights and interests states may invoke.

The first is internal to criminal law. In ascribing responsibility for actions, criminal law relies on the liberal premises mentioned at the outset: We treat each other as autonomous beings, sufficiently in control of ourselves, capable of abiding by the law, and are therefore responsible for our actions. Every criminal verdict is grounded in this arrangement. In the eyes of the law, defendants are autonomous by agreement and, hence, cannot (unless exceptions apply) appeal to a lack of factual autonomy, irrespective of undisprovable doubts and strong metaphysical objections (e.g. free will). This is political, not metaphysical freedom. But if persons are treated as autonomous beings for legal purposes, they are entitled to be treated as such on the sentencing stage as well. Treating someone as autonomous implies respect for their autonomy. If penal sanctions undermine or bypass autonomy, they contradict the basis of their legitimacy. To put it differently: If the law declares the web of causal influences that form biographies and mould persons as irrelevant in light of the *fictitious* legal concept of the autonomous person, then it cannot simply harness those factors to remould the factual citizen without falling into self-inconsistencies. Surely one may reject the arrangement and, thereby, criminal responsibility in its present form. But as long as the law upholds this premise and grounds personal responsibility upon it, internal coherence demands that it not be abandoned halfway through the process to the detriment of the defendant. The right to respect for autonomy is the reverse side of responsibility for autonomous actions.[67]

The second argument concerns limits of state powers. As punitive practices belong to the harshest measures available to the state, their limits derive from and often coincide with general limits of state authority. The state's right to punish depends on its democratic legitimacy. Likewise, it is a democratic decision whether or not a particular action should constitute a punishable offence. Violations of a norm can be understood as expressions of disagreement with its existence (regardless of whether offenders formulate it this way). Dissent from a norm neither undermines its validity, nor absolves offenders from blame for its violation. But the factual capacity and political freedom to disagree with a norm is a necessary democratic corollary from being subjected to it. A state can punish a person for violation of a norm only if it concedes him the opportunity to call it into question and repel it through democratic procedures.[68] Interventions stifling or suppressing dissenting thoughts or rebellious emotions contravene this at a very basal level. Finally, the legitimacy of punishment depends on the legitimacy of government, which in turn is grounded in the will of the people. The roots of penal legitimacy lie in the minds of the people who confer power onto the criminal justice system.

[67] That criminal law has to respect offenders' autonomy is argued in different ways by many, e.g. Duff, 'Communication', pp. 127 ff.; Morris, 'Respect'.

[68] For this reason, the disenfranchisement of felons is unjustifiable. Why should they be bound by democratic authority if excluded from democratic participation?

Conferring legitimacy is a one-directional relation from citizen onto government. If governments – through the criminal justice system – manipulate the political will of individuals and the people, they manipulate the source of their own legitimacy.[69] The free formation of moral and political opinions is, therefore, a political right and its observation a precondition of the legitimacy of state authority. Democratic punishment must leave the freedom to dissent from the norms it enforces.

These two arguments indicate how rights of offenders and limits of the state are anchored in axioms of criminal law and democratic orders. Surely, the idea of rights as constraints of public power and the promotion of overall utility may not be reconcilable under consequentialist premises. But whether a consequentialist order is preferable to a liberal democratic one is a separate matter. Here, it suffices to say that, in light of real world conditions, the development of basic human guarantees as a response to the misuses of state powers in general – and of vulnerable groups such as prisoners in particular – is promising progress, whose reversal is politically anything but desirable.

Historically, mind-altering punishments may not appear as a paradigm shift but rather as the logical continuation of the transformation of penal practices which *Foucault* famously described as moving from the body as the prime target of punishment through infliction of pain, to ostensibly more humane control of the body's outer boundaries through depriving of liberty, up to disciplining offenders through "technologies of the soul".[70] It is precisely the lack of excessive pain and the shift from corporeal to invasive yet invisible (and more pervasive) forms of control from within which characterise neuro-interventions. However, *Foucault's* description of penal practices is not in line with the evolution of the norms guiding them, clearly shying away from invasive and severe punishments: from rejecting *lex talionis* over the abandonment of corporeal punishment to the acceptance of the civil liberties of prisoners.[71] Being subjected to the harshest treatment of the state, offenders should be assured that, while they might be made to endure hard treatment, there are principled limits respectful of them as persons and of the general features of who they are.

[69] Paulo/Bublitz, 'Po(w)der to the People: Voter Manipulation, Legitimacy, and the Relevance of Moral Psychology for Democratic Theory´, Neuroethics (2016). doi: 10.1007/s12152-016-9266-7.
[70] Foucault, 'Discipline and Punish'.
[71] Hans Joas, 'Punishment and Respect', Journal of Classical Sociology 8 (2000), 159-177.