Stefano Fuselli*

Neurocorrection. On the use of neurodevices for criminals

Abstract: The possibility of using neurodevices to treat criminal offenders, as a means of voluntary diversion to avoid incarceration, has become a widely discussed topic in the last decade. A widely debated issue concerns the right to control or alter the neurological patterns of criminal offenders, provided that punishing implies limiting one's autonomy also without their consent. On the one hand, mandatory neurointervention is not only meant to be a lesser evil than incarceration, but it is even supposed to be advantageous for criminals because it can allow to restore their decisional autonomy by inhibiting their criminal impulses. On the other hand, mandatory neurointevention is rejected because it is considered to inflict significant harm on an offender, which goes far beyond the limits of criminal punishment. Some scholars have argued that the issues at stake call for a resemantisation of notions like *mental integrity*, *freedom of thought*, *and cognitive liberty*. My aim is to show that this resemantisation cannot preserve the specificity of legal categories if it is not set free from the naturalistic background, which underpins the uses of neurocorrection tools, in order to preserve the specificity of legal categories. For this purpose, I will analyse, in particular, some arguments offered by J.C. Bublitz.

Keywords: Criminal Justice; Law and Neuroscience; Mental Integrity; Neurointervention and Criminal Justice; Will.

1. A background to begin with

The amazing development of neurotechnologies has considerably increased the ways to intervene directly on and into the brain and to modify its activity. The increase is not only quantitative, but also qualitative. On the one hand, this means that there are new devices, which are by far more effective and precise. On the other hand, this implies that the spectrum of the pursued goals and purposes has become wider, thanks to the increased efficacy and functionality of the new tools.

* DPCD – University of Padova. stefano.fuselli@unipd.it.

Teoria e Critica della Regolazione Sociale, n. 1, 2021 • Mimesis Edizioni, Milano-Udine Web: www.mimesisjournals.com/ojs/index.php/tcrs • ISSN (online): 1970-5476 • ISBN: 9788857581743 • DOI: 10.7413/19705476046

Both kinds of devices have found new uses, the ones for Deep Brain Stimulation (DBS), which are implanted into the skull by surgery, and those for Transcranic Magnetic Stimulation (TMS), which do not need any surgical implant[1]. On the whole, we can speak of Electromagnetic Brain Stimulation (EBS) devices. They have turned out to be particularly useful and effective in patients who are not responsive to pharmacological treatments anymore. Additionally, their effects are quicker, limited to the targeted areas, and immediately reversible. All this has contributed to widen the range of their use from the initial treatment of PD patients or some brain dysfunctions like epilepsy to the treatment of depression or anorexia.

Their therapeutic potential has become even more attractive. Thanks to their connection to the Brain Computer Interface (BCI) technology[2], it is possible to monitor in real time the electro-brain activity and to transform it into an output signal, which can be codified and transmitted by means of software[3]. Thus, several different applications have become possible, such as implementing neuroprosthetics by means of which a patient can steer a wheelchair, or constantly monitoring the brain activity of an epileptic patient in order to detect the signals of a forthcoming attack.

The converging of technologies that send impulses *to* the brain and transform into impulses the signals coming *from* the brain has opened the way to the development of increasingly effective tools. New devices that can regulate the intensity of the stimulation according to the actual condition of the patient can also anticipate the condition of the patient and activate or regulate the stimulation *before* an event happens. Moreover, there are devices that can modulate the stimulation automatically by means of software, totally bypassing the control of the implanted subject.

2. A new frontier for DBS

It is worth mentioning that the possibility of using totally automatized DBS devices has sparked off an intense debate on their compatibility with individual autonomy. Many scholars have claimed that an external activity, which can influence the neurological bases of the mind functions that underpin actions, thoughts, and emotions without the interested person having any control on the sources of this activity, is a threat to personal autonomy, independently from any therapeutic usefulness. By contrast, many others have argued that recuperating certain neurological functions by means of self-regulating devices restores personal autonomy, both because patients are relieved from the boundaries of their pathology and

1 See, for an introduction to this topic Reti and Chang 2015.

2 See, for a recent outlook on this topic, Guger et al. 2019.

3 The literature on the topic is overwhelming; for a first approach, see Folgieri 2020. It is worth mentioning that scientists are working on an integration of biological and artificial synapsis, which makes it possible to detect and process neuronal spikes even from thousands of miles away; see Serb et al. 2020.

because the stimulation works only on processes that usually happen behind consciousness, even in the case of healthy people[4].

Another use of DBS is particularly relevant for the issue at stake in this paper. Some years ago, an Italian equipe has experimented it in subjects suffering from mental retardation with aggressive and disruptive behaviour and resistant to any pharmacological treatment. The DBS of a targeted small area (the posteromedial hypothalamus) had quick (and reversible) positive effects. After a while, some patients could be discharged from the institution where they had been hospitalized and began to attend a psychiatric rehabilitation centre[5].

These results have opened up a new perspective in the already intense debate on the possible and permissible uses of neurotechnologies to tackle crime. Provided that these technologies can both stimulate the brain and monitor or even predict brain activity, the issue at stake is whether they could be used for criminal offenders both to treat them and to prevent them from committing crimes and to what extent it would be permissible[6].

Of course, "currently much of the support for the use of EBS to reduce or prevent criminal behaviour is highly speculative, consisting mainly of small-scale and as-yet unreproduced cognitive science experiments"[7]. Nevertheless, because of the technological developments, this kind of intervention cannot be relegated to a distant dystopian future. Thus, taking the discussion seriously is not unreasonable, even if one of the most frequently used words in this debate is "suppose"[8].

3. Advantages

The issues in question entangle different levels. To begin with, one problematic aspect is provided by the criteria to select who is authorised to implant those devices and to establish whether the devices need to be completely self-regulating or if they should simply signal to the implanted subject an abnormal neural activity, prodromal of violent behaviour[9].

Nevertheless, the use of DBS devices seems a good way to overcome the objections against direct brain intervention on criminal offenders. Brain activity can be directly influenced not only by acting on the electric impulses, but also by means of drugs, which alter the biochemical neurotransmission processes. From the daily use of caffeine to the pharmacological treatment of depression, there is a wide range of substances, which are used for therapy or stimulation of brain activity.

4 For an overview of the literature on the main ethical and legal issues at stake, as well as for a first philosophical outline, see Fuselli 2020.

5 Franzini et al. 2005; Franzini et al. 2013.

6 Among the first scholars who discussed the topic on the basis of this Italian study, Greely 2008; Greely 2009. On the use of different devices from those for DBS, see McMillan 2018: 231–32.

7 Chew et al. 2018: 32.

8 McMahan 2018.

9 Gilbert 2015.

Currently, some drugs like medroxyprogesterone acetate and cyproterone acetate are used for what is known as 'chemical castration' in the USA and some European countries, respectively[10]. They greatly lower the testosterone level in males, causing a sharp decline in sexual impulses in general. Thus, they act in a non-selective way, inhibiting not only –say– paedophilic sexual impulses, but also 'licit' sexual impulses of the treated subject. That is to say, not only do they bypass every control of the treated person, but their effects also exceed the extent of the crime one has been sentenced for. Direct brain intervention by means of drugs interferes also with the mental states of the criminal offender that are different from those which are possibly targeted for treatment. In addition, they have a lot of side effects, which can be severely dangerous for the health of the subject. In any case, they have consequences on aspects that are not censurable and are not connected with the crime that the offender was sentenced for[11].

This is one reason why predictive brain-implants are meant to be able to get rid of the drawbacks of using drugs. Indeed, if they can be programmed to detect and reveal neural activity prodromal of aggressive behaviour, they could also be put under the control of the implanted subjects. In this way, the subjects could choose whether to activate the inhibition stimuli and so they would be wholly accountable for their behaviour[12], a kind of induced conscience, one might say[13].

4. Equivalence and coercion

To sum up, neurodevices are viable candidates as ideal tools for direct intervention on criminal offenders' brain activity. Compared with pharmacological tools already in use, they seem to act more selectively, precisely and effectively; moreover, they are completely reversible. Additionally, due to the highly sophisticated technology they are based on, they could also provide a kind of warning signal to the implanted subject, without bypassing their conscious control and the possibility of freely choosing how to behave. Thus, their function can be twofold, not only corrective or rehabilitative, but also predictive and preventive.

Apart from the concern or enthusiasm raised by the possible use of these devices, the function they would have in criminal justice has been one of the first critical issues to deal with. From this point of view, considering that treatment has turned out to be the most supported option, the crux is whether neurointervention should be only voluntary or even mandatory. In the latter case, the question is whether it should be a part of the deserved punishment (if not a form of it)[14].

Due to the specificity of this technology, the more general debate on neurointervention for the treatment or punishment of criminal offenders provides the general

10 Greely 2009; Douglas et al. 2013; Chew et al. 2018; Forsberg 2018.
11 Birks and Buyx 2018; Buyx and Birks 2018.
12 Ryberg 2015.
13 On the relationship with the topic of moral enhancement, see Wiseman 2016.
14 On the relevance of the cultural background of this debate, see Matravers 2018.

background also for this particular topic. At least two issues of the wider debate on neuro-correction are worth mentioning here.

The first one is the so-called equivalence thesis, according to which both traditional and new instruments aim to change the offender's behaviour. The claim is that there is no substantial difference between acting directly on the brain of a criminal offender by means of drugs or devices and, indirectly, by means of traditional instruments like cognitive therapy or incarceration. Given that the goal of criminal justice is not only retribution, but also achievement of something useful for the society and the offender, also via rehabilitation or prevention of criminal behaviour, if direct brain interventions are proved to be sufficiently effective with an acceptable risk level, there should be no more reason to reject them[15]. Additionally, some scholars claim that these instruments might be not only more effective, but also even less threatening for the offender's autonomy than incarceration, or at least no more threatening than other medical interventions[16].

Directly connected to the equivalence thesis, the second issue is whether direct brain intervention for criminal offenders should be mandatory or only voluntary[17]. Generally speaking, the claim that direct brain intervention may be non-consensual is connected with the idea that crime –especially in the case of violent or sexual crimes– is a consequence of deviance, a sign of dysfunction that the state has the right and duty to neutralize and possibly to correct or prevent, in the same way that it would act to protect people in case of an epidemic through compulsory vaccination or quarantine[18]. Additionally, if one admits that the offender's best interest should not necessarily set a limit for neurotechnological treatment[19], mandatory neurointervention is even more permissible in the case of psychopaths or mentally insane offenders[20].

Both these issues, equivalence and coercion, are at stake also in the discussion concerning monitoring and regulating brain electro activity by means of neurodevices. The examination of connected features, like their being more accurate and selective, on the one hand, and the additional need for a surgical implant, on the other, is generally absorbed into the topics mentioned above.

5. Some concerns

Challenging the equivalence thesis, many scholars have objected that there is an unbridgeable difference between the interventions that bypass the mental control

15 Greely 2008: 1134.

16 Douglas et al. 2013; Ryberg and Petersen 2013; Pugh and Douglas 2016; Shniderman and Solberg 2018.

17 Notice that according to Greely, they can be used only if they are accepted "voluntarily by an informed, competent adult", Greely 2008: 1134.

18 Douglas et al. 2013; Pugh and Douglas 2016; Petersen and Kragh 2017.

19 Petersen 2018.

20 Lavazza 2018a; Palk 2018.

capacity of the receivers and those that permit or even request that the receivers be fully aware of what they are going through[21]. The distance does not depend on quantitative features –like, f.i., being more or less effective, more or less reliable– but on qualitative traits connected to the conditions and the goal of their action.

Indeed, instruments that act directly on the brain make any kind of resistance impossible[22]. Additionally, they operate in a different way on the cognitive system. The usual stimuli which we are affected by, no matter whether we are aware of them or not, are inputs that are processed by our "cognitive machinery"[23] according to its predispositions. By contrast, direct interventions aim to change the cognitive machinery itself, in order to modify the outputs.

Thus, some scholars have argued that direct interventions undermine our sovereignty over our mind, because they manipulate the neuronal correlates of our mental functions without any possibility to resist. Therefore, they should be declared illegal and punishable as a criminal offence, since they also undermine the human right to freedom of thought[24].

On the other hand, some scholars do not object to the use of direct brain intervention for criminal offenders as a matter of principle, but only against mandatory interventions. Mandatory neurocorrectives should be rejected as a practice that diminishes the receivers to a sub-human level because it treats them as objects and harms their mental and bodily integrity. By contrast, if offenders are given full information about the effects and efficacy of a treatment by means of direct brain intervention and their consent is genuinely voluntary and informed, the state could be legitimised to offer this possibility as an alternative to incarceration or as a condition for early release[25].

Of course, supporters of this claim have to establish the conditions for consent to be valid when given by individuals who are imprisoned[26]. Additionally, consent to direct brain intervention does not eliminate the inconsistency of the goal. Another issue they have to tackle is the consistency of the notion of rehabilitation as something done to the offender rather than by the offender with the goal that the treated subjects possibly become autonomous and not automatons[27].

Furthermore, the use of DBS devices in psychopath detainees raises questions connected to medical ethics and practice. In a recent paper, the legitimacy of carrying out experiments on psychopaths has been challenged. The claim is that there are no conditions for a valid informed consent and there are no guarantees of

21 The distinction was already driven by Levy 2007: 70. This is the reason why Shaw 2014; Shaw 2018 challenges the rehabilitative function of neurointerventions.

22 Bomann-Larsen 2013.

23 Bublitz and Merkel 2014; Bublitz 2015.

24 Bublitz 2014; Bublitz 2015; Bublitz 2016. For a possible objection see Ryberg 2015, according to whom neurodevices can be programmed to detect and reveal neural activity prodromal of aggressive behaviour, instead of acting directly on the brain. It is my view that the issue at stake is of a different kind, as I aim to demonstrate.

25 Shaw 2015.

26 Bomann-Larsen 2013; Shaw 2015.

27 Kirchmair 2019.

any therapeutic advantage, provided that some features of psychopathic personality are the absence of subjective suffering and the lack of moral motivation[28]. Furthermore, the doubtful therapeutic efficacy of DBS treatment for psychopaths provides an *a fortiori* argument, although empirical and contingent, to reject the possibility of it being mandatory. If the DBS of some regions associated with psychopathology can guarantee neither the social integration of psychopaths nor their becoming non-psychopaths, then it is not clear who should be in charge of the decision, nor which should be the parameters for the intervention, nor which is the aimed final setting of personality alteration[29].

6. A comeback: the deviant

The promotion of mandatory brain intervention involves a particular theory of punishment, whose pivotal assumption is that the offender, especially in the case of violent or sexual crimes, is a deviant, that is, an abnormal and dangerous individual. The aim of intervening on a personality in order to modify it, regardless of whether this is meant to be corrective or rehabilitative, is not necessarily to produce some beneficial effects on the individual, but to protect the society like in a vaccination campaign[30].

Far from being new, this theoretical approach to criminal justice draws fresh nourishment from neuroscience and, especially, from a general mind-set, according to which mental and behavioural phenomena and the constitutive traits of personality are direct expressions of brain activity, which are more or less localisable in specific brain areas. Committing a crime counts as a manifestation of a dysfunction in the brain activity of the offender, who needs to be treated in order to come back to normality or, at least, to become inoffensive for the society[31]. Therefore, there is no clear cut between punishment and treatment, and direct brain intervention could be mandatory or at least consensual.

The offender's personality, rather than the committed crime or the harm caused to the victim, has an undeniable pivotal role in the proposed neurocorrectives or in the reasons by means of which they are promoted. The attempts to justify them from a retributivist point of view, like by arguing that they could purposely be designed to inflict pain in order to satisfy the requisite for punishment to be afflictive, are "particularly regrettable"[32]. Moreover, they are clues to an oversimplified, reductive, and even misrepresented theory of punishment. In order to tackle the claim that the use of such instruments might count for punishment, referring to Article 3 of the European Convention of Human Rights[33] and to the statements

28 Hübner and White 2016.
29 Mackenzie 2016. For a different opinion, see Ryberg 2016.
30 Pugh and Douglas 2016.
31 On normality and normalisation by means of neurocorretives, see Sommaggio 2016.
32 Kirchmair 2019: 25.
33 "No one shall be subjected to torture or to inhuman or degrading treatment or punish-

of the European Court of Human Rights is a considerably remarkable move[34], at least in the European context. Indeed, the Court stated that a treatment or a punishment is inhuman and degrading also when it causes the "breaking of an individual's moral and physical resistance"[35].

7. *Repetita nocent*

The idea of using neurocorrectives, and in particular neurodevices, has consequences that go beyond the grounds of the current debate. However unpleasant or questionable it may sound, it has the merit to stimulate a more wide-ranging reflection.

As we have seen above, the new opened front calls for protection of aspects, which the traditional categories of law are not fit for and criminal law has to equip itself properly for. Many scholars have repeatedly warned about the need for a resemantisation of notions like *mental integrity* [36], *freedom of thought, cognitive liberty*[37], because of the findings and the related potentialities and risks of neurotechnologies. The plea to extend the range of human rights or to reinterpret them in the light of the new challenges is perhaps the clearest expression of that exigency.

In scientific and technological research, every crucial turn that produces an increase of power causes also a loss of innocence. The world, in which the law used to regulate only the exterior behaviour, might be forever compromised or even lost, because the distinction itself between interior and exterior is gradually becoming more and more nonsensical. In the meanwhile, focusing on some funding traits of this process and on the categories into which they are usually moulded, as well as on the ways of facing them becomes increasingly urgent. Understanding what is going on is the precondition to develop a set of new fitting tools to govern these dynamics.

In order to clarify the kind of intellectual enterprise requested by neurotechnology, I will discuss one of the theses mentioned above, analysing its conceptual underpinnings and outlining the background that needs to be considered. The argument has been presented and developed by Bublitz in several contributions. He holds a clear-cut position both *against* the use of neurocorrectives and *for* a deep revision of the fundamental rights of freedom and integrity established by different Charts. My aim is not to challenge these theses, because I agree with the concerns that nourish them, but rather to analyse some terms and assumptions by means of which he supports them.

ment", ECHR, Art. 3.

34 Shaw 2018; Kirchmair 2019.

35 For this quotation, see Kirchmair 2019: 30.

36 Ienca and Andorno 2017; Lavazza 2018b.

37 Sententia 2004; Bublitz 2014; Bublitz 2015; Bublitz 2016; Craig 2016; Bublitz and Merkel 2014.

To begin with, the difference between indirect and direct brain interventions provides one main reason to challenge the equivalence thesis. According to Bublitz and Merkel, direct brain interventions not only bypass any conscious control on the part of the receiver, but they also act differently from the indirect interventions, because "*indirect interventions are inputs into the cognitive machinery our minds are adapted to process, whereas direct interventions change the cognitive machinery itself*"[38]. However, this clear-cut distinction between indirect and direct brain interventions fades and becomes less stable, when Bublitz states that "[d]irect stimulation of the brain and persuasive arguments might be conceived as two poles in a broad spectrum of gradually different interventions"[39].

It is my view that the only possibility for characterising both direct brain stimulation (via electrodes, magnets, drugs) and the use of persuasive arguments as two poles of the same spectrum –or, in a more traditional lexicon, as two extremes of the same genus– is provided precisely by the metaphor of the *machinery*. Indeed, it appears not to be neutral, when one considers the kind of relationship it affords. Generally speaking, machinery is nothing but a bare passive receiver of the relationship the acting subject has with it. In this connection, there is no kind of reciprocity. Moreover, machinery does not activate itself spontaneously, because it needs external impulses, and it only gives outputs after it has been induced to do so by a type of external input. It is a one-sided relationship, without any sort of co-protagonist, because on the other side there is no subject.

Qualifying the communicational relationship which underpins persuasive activities as a relationship to a machinery, although only metaphorically, leads to forever forsaking one of the possible meanings of mental integrity and freedom of thought, that is the meaning which is specific for intersubjective relationships. One could say that all types of manipulative relationships, no matter whether direct or indirect, deny intersubjectivity impeding cooperation, because on the one side there is nothing but a bare receiver, which has no reason to be there but in terms of the goals or outputs that the source of the inputs aims at. Yet, Bublitz seems to recognise this in admitting that the difference between the interventions designed to bypass consciousness and the other ones draws a distinction line in the spectrum[40]. What the non-bypassing interventions cater for is precisely what the first ones deny, namely, that one term of the relationship is irreducible to a bare receiver of the other one's action.

Far from being simply pedantic, this remark has an immediate consequence on the idea of crime and punishment. Crudely speaking, a crime is mainly a crime because it impedes all kind of cooperation. It harms, foremostly, the victim's sub-

38 Bublitz and Merkel 2014: 70.

39 Bublitz 2015: 1325.

40 "As a first approximation, a concededly rough line can be drawn between interventions that intentionally bypass control capacities or exploit cognitive weaknesses on the one side and interventions that, at least in principle, respect control and freedom of thought as they do not undermine powers of resistance on the other. One example is subliminal messages entering minds through the senses without rising to conscious awareness. Because they are designed to bypass conscious control, they regularly do not respect freedom of thought of the receiver", Ibid.: 1325.

jectivity, rather than their material goods or even their bodily integrity, because the criminal action reduces the victim to a passive receiver. At the same time, however, punishment is not allowed to transform a criminal into a passive receiver. If one accepts, following the European Charts, that also the offender's mental integrity deserves protection, what acts in the same way as the offender acted cannot provide any form of punishment or be any part of the punishment at all.

Thus, it is my view that the use of direct brain interventions via drugs or devices cannot be otherwise justified but for therapy, that is, for rehabilitation, if possible, of the neurological condition that enables the patient to have cooperative relationships. Inasmuch as it is therapeutic, such a use is properly not a form of or a part of punishment, nor a form of diversion from incarceration.

At the same time, I argue that they cannot even be seen as part of a rehabilitation programme in a properly juridical sense. Let us consider, for example, the Italian Constitution, according to which "[p]unishments may not be inhuman and shall aim at re-educating the convicted" (Art. 27, c. 3). Clearly, re-education is not rehabilitation in a medical, but in a social sense. Namely, the convicted should be re-enabled to a different kind of relationship from the one he/she enacted through the crime. If it is questionable whether incarceration is a fitting instrument to achieve this goal, any kind of intervention that reiterates the same form of relationship enacted in the crime is *a fortiori* unfit. Rehabilitation is not something done *to* the offender but *by* the offender.

8. On the notion of mental integrity

In a well-known paper, published at the beginning of this century, Green and Cohen argued that, for the law, neuroscience would change nothing and everything, in particular in criminal justice[41].

Neurocorrectives, of which the ones using brain electro activity stimulation are a particular type, could be seen as a proof that the two scholars had a keen insight. As we have seen above, people who promote and support these techniques do not aim at abolishing punishment, but rather at making some of its functions more efficient, in particular prevention, rehabilitation, social security, and the causing of a certain degree of deserved suffering.

Apparently, nothing changes. Substantially, everything does. The aim of the cyclopic collective enterprise called neurolaw is not to check the compatibility of juridical categories with neuroscientific conceptual tools[42], but to adapt juridical cat-

41 Greene and Cohen 2004.

42 "Neurolaw is a descriptive and normative field in which scientists and legal scholars seek to apply recent advances in genetics and neuroscience to the classical conceptions of law, with the aim of verifying whether legal institutions are consistent with current scientific knowledge and, if not, proposing changes to that effect (cfr. Pardo and Patterson 2014). It is a process of naturalization *à la* Quine, which so far has not affected the law", Lavazza 2018a: 154.

egories into the neuroscientific categorical frame[43]. In the light of neuroscientific advances and of the naturalization of mind processes, which is specific to this kind of research field, juridical categories are naturalized too, in particular in criminal justice. Thus, rehabilitation, protection, and prevention are becoming synonyms for monitoring and inhibiting violent and aggressive impulses or weakening and diminishing sexual libido.

As a consequence, every attempt to understand action and behaviour according to other conceptual modalities becomes impossible, because nonsensical. There is no more room for a different perspective, which would allow to shape specific conceptual tools for the understanding and regulation of the relationships every society is woven by. Not accidentally, a scholar has recently invoked an "epistemic abstinence"[44] towards the modern model of approaching reality and its claim to be the unique possible one. While acknowledging the usefulness of this limiting epistemic caution, it is my view that the issue at stake needs also a constructive effort to explore how and to what extent the different components of the present debate could be harmonized.

From this point of view, the claim about the need to redefine mental integrity[45] or even to introduce the notion of cognitive liberty could provide sufficient evidence that this path is going to be difficult[46]. On the one side, this claim reflects awareness of the fact that advance in neuroscience has opened up new fronts, which were unimaginable until now, because it seems to have become possible to have effective targeted control over the processes that underpin volition and individual behaviour. On the other side, it seems not only to acknowledge that the scenario has changed, but rather to legitimate the naturalistic outlook, according to which there is no possible way to understand will and self-determination beyond what is made visible and manipulable by the new categories and tools.

Understanding mental integrity as the autonomy, mastery, or sovereignty[47] that people have on their own mental states is not sufficient to outline a different, non-naturalistic way of being of the will and its determinations. If the only possible

43 Fuselli 2016; Nunziante 2016.

44 Bennett 2018: 256.

45 Ienca and Andorno 2017. For instance, *mental integrity* is defined as "the individual's mastery of his mental states and his brain data so that, without his consent, no one can read, spread, or alter such states and data in order to condition the individual in any way", Lavazza 2018b: 4.

46 "Cognitive liberty is a term that updates notions of 'freedom of thought' for the 21st century by taking into account the power we now have […] to monitor and manipulate cognitive function. Cognitive liberty is every person's fundamental right to think independently, to use the full spectrum of his or her mind, and to have autonomy over his or her own brain chemistry", Sententia 2004: 222–23. For a discussion of the notion of *cognitive liberty*, with particular regard to Bublitz, see Sommaggio et al. 2017.

47 "Second, in light of a right to mental self-determination, interventions bypassing mental control are evidently illegitimate. Most indirect interventions, by contrast, at least in principle obey the command to respect the other's *mental sovereignty*. In consequence, we may hurt people with cutting remarks and manipulate them with distortive words, but not by directly tinkering with their neurons", Bublitz and Merkel 2014: 73. "If freedom of thought demands

conceptual background is the neuroscientific one, then autonomy, mastery, or sovereignty themselves are nothing but reflexes or expressions of an underpinning manipulable neurological pathway, dynamics, or state. At the end of the day, if nobody is impervious to the environment –no matter if natural or social– and nobody can assume to be the unique source of their own mental states, the notion of a right to master them becomes rather uncertain[48].

Differently, one can approach the issue from another perspective, in which the will is not reducible to the neurological or mental activity of an individual, even if it is not disjointed from it. Should this be possible, then a meaning of mental integrity is also possible, which requires the existence of an asset of specific categories, but is not indifferent to the phenomena studied by neuroscience.

In the next section, I am going to illustrate this possibility by means of two different categories of examples, one based on the field of neuroscientific research, and the other provided by the history of political thought. The aim is to outline some forms of the individual will that go beyond both the neurological and the mental state levels.

9. Forms of the will

As paradoxical as it may sound, two leading neuroscientists, Benjamin Libet and Gerald Edelman, were among the first ones who argued that there are some ways of being of the will, which do not depend on the underpinning neurological activities of an individual.

In the current debate, the notion of *veto* is a well-known and widely discussed one. Benjamin Libet uses the word *veto* to denote the conscious act by means of which the process that leads to 'act now' is blocked. According to Libet, the process leading to a voluntary act is initiated by the brain about 400 msec before the conscious will to act 'now' appears. After the subject becomes "consciously *aware* of the urge or wish to act"[49], the actual movement of the muscle activation takes another 150 msec, during the last 50 of which "the act goes to completion with no possibility of its being stopped"[50]. In the remaining 100 msec, the consciously aware subject could intervene, deciding whether to allow the process to go to completion or to veto it, so that no action occurs.

Veto is a very particular function, because Libet did not find any trace of neuronal activity connected to a spontaneously vetoed intention to act[51]. Thus, he proposes that the conscious veto may not require or directly depend on preceding unconscious processes. The reason he provides is not only empirical, but also theo-

that we respect each other's *mental sovereignty* and forbids manipulating thinking processes (…)", Bublitz 2015: 1325. My italics.

48 Bennett 2018: 260.

49 Libet 2004: 134.

50 Ibid.: 138.

51 Ibid.: 141–42.

retical: "The conscious veto is a control function, different from simply becoming aware of the wish to act. There is no logical imperative in any mind-brain theory, even in identity theory that requires specific neural activity to precede and determine the nature of a conscious control function"[52].

The role of conscious veto is of the highest importance, even if the action is performed. Indeed, it provides a possible discontinuity in the linear process, which goes from cortex activity to muscle activation. Thus, the *reality* of a neural process becomes the *possibility* of two different solutions, and "what is lost in terms of *linearity* is gained in terms of *complexity*"[53]. Therefore, a margin remains – apparently very narrow, although really decisive – for a difference coming into being, not only on the ontological level, but also on the practical one. A space is opened up here for a kind of action that is not the effect of a bottom-up causation process.

The claim of Gerald Edelman, the author of *Neural Darwinism*[54], is perhaps less well known than Libet's, although he won the Nobel prize. In his book *Bright Air, Brillant Fire. On the Matter of the Mind*[55], he argues that at the base of the symbolic activities that are typical for humans, among which language is by far the most relevant, there is a higher-order consciousness, which is different from the primary consciousness.

The latter manages the aversive or appetitive reactions of an organism to the environment depending on its homeostatic values and the salience of external signals in terms of danger or reward[56]. The limit of this kind of consciousness is that, on the one hand, the organism is confined to a sort of eternal present, because it "does not afford the ability to model the past or the future as part of a correlated scene" and, on the other hand, the organism "lacks an explicit notion or a concept of a personal self"[57].

52 Ibid.: 147. Various experiments have been conducted to find the neural correlates of veto; see f.i.: Brass and Haggard 2007; Walsh et al. 2010; Filevich et al. 2013. Beyond any experimental finding, it is worth quoting Libet: "The possibility is not excluded that factors on which the decision to veto is based do develop by unconscious processes that precede the veto (…) The awareness of the decision to veto could require preceding unconscious processes, but the content of that awareness (the actual decision to veto) is a separate feature that need not have the same requirement", Libet 2004: 147. From this point of view, it may be unsafe to conclude that "[t]here can be no changes on the mental level without some change on the physical, i.e. neuronal level. Or, to put it slightly differently: mental states not only correlate with particular brain states, but are also "caused" or (somehow differently) "realized" by physical states ("bottom-up" causation)", Bublitz and Merkel 2014: 54. Consequently, the claim that "[t]he law should discard dualistic differentiations, replace them with the assumption that all mental phenomena supervene on (or at least correlate with) neuronal processes and begin to develop a theory of which mental phenomena ought to be protected on this premise", ibid.: 55, could undoubtedly play a heuristic and precautionary role, but not outline any definitive horizon.

53 Chiereghin 2008: 308–09. My italics.

54 Edelman 1987.

55 Edelman 1992.

56 Ibid.: 121.

57 Ibid.: 122.

By contrast, having a high-order consciousness means to be "conscious to be conscious"[58]. Among the various abilities that a self-conscious organism affords, there is the capacity of deferring the satisfaction of impulses or even of denying it completely, because creatures endowed with high-order consciousness can "transcend the dictates of biology"[59]. According to Edelman, the arising of high-order consciousness has provided human organisms with the possibility not only of modifying values at a biological level, but foremostly of performing a "*total* denial of biological values", at least "on the part of those organisms we call martyrs and saints"[60].

Overestimating the extent of that *total* is an arduous task, particularly when the focus is on the neural activities and processes that underpin the flourishing of mind potentialities. High-order consciousness enables mental phenomena that are not totally dependent on the neurobiological dynamics and values, which they are rooted into, to such an extent that Edelman himself defines the change induced by the arising of this kind of consciousness as an "ontological revolution"[61].

To sum up, according to Libet and Edelman, in the processes where the will does not perform any executive act, but is nevertheless fully effective on reality to the extent of giving it an ontological turn, it seems neither to leave any tracks in the neural processes that usually underpin one's activity, nor to be entirely conditioned by them.

This abstaining form of the will plays a pivotal role also in the philosophical thinking on sovereignty, a notion that we have seen being used in the contemporary debate on mental integrity.

Among the various theories formulated during the past centuries, the one provided by Rousseau is worth mentioning, with particular regard to the kind of will characteristic of the sovereign political body, that is, the *general will*. Indeed, the general will has nothing to do with a mental state, because it is neither an individual's will, nor the will of all[62].

The general will arises because each associate, with his individual will, gives himself up entirely through the total alienation of himself as well as his rights and interests to the whole community[63]. By performing this act, each single person shows his capacity of transcending subjective individual will with its connected needs. The consequence is that,

> in place of the particular individuality of each contracting party, this act of association produces a moral and collective body, composed of as many members as the assembly has voices, and which receives from this same act its unity, its common self (*moi*), its life, and its will[64].

58 Ibid.: 131.
59 Ibid.: 163.
60 Ibid.: 163. My italics.
61 Ibid.: 150.
62 See Rousseau 2002: 172.
63 Ibid.: 163.
64 Ibid.: 164.

The possibility of thinking of an a-subjective or un-mental dimension of the will as the proper body of the political community –therefore also of that kind of will which makes laws– does not mean that there are no real individual subjects in the collective body. Rather, their subjective psychical processes are not enough to understand and to explain the meaning of the acts, which are specific of the different kind of reality produced by the social contract.

10. So what? A provisional conclusion

Indeed, the examples mentioned above are quite different from an epistemological point of view. The first two come from empirical research, whereas the last one comes from political philosophy. Nevertheless, they show that the will affords to reach a different level compared to the underpinning phenomenal one (the one of electric impulses and chemical reaction and the one of subjective individual psyche, respectively). Being irreducible to that phenomenal level without being separable from it is the way in which this capacity reveals itself and becomes effective.

It is my view that advances and findings in neuroscience and neurotechnology call for constantly revising juridical tools. In the meanwhile, what distinguishes law from other fields and makes its conceptual asset specific cannot be overlooked, on pain of giving up law itself.

Legal *fictions* are tools the law has shaped in order to cope with forms of the will that are not reducible to a subjective mental state. Let us consider, for instance, the idea of the legislator's will, which, in the continental legal tradition, denotes the intention of the legislative body (such as Parliament) taken as a whole and is used for the interpretation of statutes[65]. Clearly, it is a kind of will that has nothing to do with the underpinning psychical dynamics or neural processes, but the statute is considered a manifestation and product of a will[66]. From this point of view, the choice itself of protecting certain rights, among which the right to mental integrity, is an expression of a will that transcends the level of mental phenomena or states of mind.

The idea that liability could be attributed also to a non-psychical will is not unknown to the criminal law itself. For example, in Italy criminal responsibility has been extended also to legal persons or organisations[67], partially making an

65 See: *volontà del legislatore* (Italian); *volunté du législateur* (French); *Wille des Gesetzgebers* (German); *voluntad del legislator* (Spanish).

66 One can argue that it is nothing but a figure of speech. As a possible objection, also saying that 'the brain thinks' is a figure of speech, and precisely a synecdoche if not a mereological fallacy; on the topic, see Bennett and Hacker 2003: 68–107. The non-psychical dimension of the will is maintained also when the notion of sovereignty is replaced with the one of autonomy, the latter being meant as "setting ends for oneself, to be independent of another's will, and in Kantian terms, to be a moral self-legislator", Bublitz 2018: 302. According to Kant, the possibility for the reason to be practical by itself, that is, to enact autonomous will, is provided by the categorical imperative of the moral law, which has no personality.

67 Legislative Decree No 231/2001.

exception to the principle *societas delinquere non potest*. On a different level, the non-psychical or mental aspect of the will allows punishment to be different from a subjective reaction to violence and crime, and to have a public and not private character, like revenge.

Of course, the naturalistic, psychical, and mental traits of will are fundamental for law, in particular for the criminal law: a kind of *iustitia*, self-referentially enclosed in a perfect mechanism of *fictions*, which does not cope with *mundus* makes no sense. The naturalistic traits have a pivotal role with regard to the *mens rea*, the psychological elements of crime, the liability of the offender, and the existence of a crime. Nevertheless, these dimensions do not cover all the meanings of the related legal notions, which have been shaped to understand those aspects of reality, like behaviour or intersubjective relationships, whose meaning goes beyond the neuroscientific research field[68]. Indeed, one could say that the first and fundamental legal fiction is the notion of autonomous person[69].

Moving on, the notion of mental integrity should cope with the complex nature of will, whose manifestations allow to transcend the respective levels of their condition of possibility. If will is admissibly not reducible to the recorded neural activity data, protecting mental integrity is also and foremostly protecting the irreducibility of the subjective will –including that of the criminal offender– to a bare epiphenomenon of neural activity, which is precisely the conceptual background of neurocorrection. In addition, if will is admissibly not only a subjective mental state or function, the protection of mental integrity reaches a further dimension, which underpins the fact of living in an organized community. In this perspective, protecting mental integrity implies admitting that will makes it possible to transcend the individual psychical level and to achieve other ways of existing.

To sum up, neurocorrectives are an unjustifiable and unacceptable harm to mental integrity because they overlook its complexity and downgrade it to a level, which cannot provide a sufficient account of the multifaceted manifestations of an individual's will. Whatever neurocorrectives may promise, it cannot be in the name of criminal justice.

References

Bennett C. 2018, "Intrusive Intervention and Opacity Respect", in: D. Birks and T. Douglas (eds) 1995, *Treatment for Crime. Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford University Press: Oxford, pp 255–73.

Bennett M.R. and P.M.S. Hacker 2003, *Philosophical Foundations of Neuroscience*, Malden (Mass): Blackwell.

68 On the topic, Morse 2011a; Morse 2011b; Pardo and Dennis Patterson 2013.
69 Cfr. Bublitz 2018: 316.

Birks D. and A. Buyx 2018, "Punishing Intentions and Neurointerventions", *AJOB Neuroscience*, 9:133–143.

Bomann-Larsen L. 2013, "Voluntary Rehabilitation? On Neurotechnological Behavioural Treatment, Valid Consent and (In)appropriate Offers", *Neuroethics*, 6: 65–77.

Brass M. and P. Haggard 2007, "To Do or Not to Do: The Neural Signature of Self-Control", *Journal of Neuroscience*, 27: 9141–9145.

Bublitz J.C. 2015, "Cognitive Liberty or the International Human Right to Freedom of Thought", in J. Clausen and N. Levy (eds) 2015, *Handbook of Neuroethics*, Dordrecht: Springer Netherlands: 1309–1333.

Bublitz J.C. 2014, "Freedom of Thought in the Age of Neuroscience", *Archiv für Rechts- und Sozialphilosphie*, 100:1–25.

Bublitz J.C. 2016, "Moral Enhancement and Mental Freedom", *Journal of Applied Philosophy*, 33:88–106.

Bublitz J.C. 2018, "'The Soul is the Prison of the Body' – Mandatory Moral Enhancement, Punishment & Rights Against Neuro-Rehabilitation", in D. Birks and T. Douglas (eds) 2018, *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford: Oxford University Press: 289–320.

Bublitz J.C. and R. Merkel 2014, "Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination", *Criminal Law and Philosophy*, 8:51–77.

Buyx A. and D. Birks 2018, "Neuroscience and Social Problems: The Case of Neuropunishment", *Cambridge Quarterly of Healthcare Ethics*, 27: 628–634.

Chew C., Douglas T. and N.S.Faber 2018, "Biological Interventions for Crime Prevention", in D. Birks and T. Douglas (eds) 2018, *Treatment for Crime. Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford: Oxford University Press: 11–43.

Chiereghin F. 2008, "La coscienza: un ritardato mentale?", *Verifiche*, 37: 283–316.

Craig J.N. 2016, "Incarceration, Direct Brain Intervention, and the Right to Mental Integrity – a Reply to Thomas Douglas", *Neuroethics*, 9:107–118.

Douglas T., Bonte P., Focquaert F., et al. 2013, "Coercion, Incarceration, and Chemical Castration: An Argument From Autonomy", *Journal of Bioethical Inquiry*, 10: 393–405.

Edelman G.M. 1987, *Neural Darwinism: The theory of neuronal group selection*, New York: Basic books.

Edelman G.M. 1992, *Bright Air, Brillant Fire. On the Matter of the Mind*, New York: Basic books.

Filevich E., Kühn S., and Haggard P. 2013, "There is no free won't: Antecedent brain activity predicts decisions to inhibit", *PloS one*, 8(2), e53053.

Folgieri R. 2020, "Brain Computer Interface and Transcranial Magnetic Stimulation in Legal Practice and Regulations", in A. D'Aloia and M.C. Errigo (eds) 2020, *Neuroscience and Law. Complicated Crossings and New Perspectives*, Switzerland: Springer: 273–90.

Forsberg L. 2018, "Crime-Preveting Neurointerventions and the Law. Learning from Anti-Libidinal Interventions", in D. Birks and T. Douglas (eds) 2018, *Treatment for Crime. Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford: Oxford University Press: 44–70.

Franzini A., Broggi G., Cordella R., et al. 2013, "Deep-Brain Stimulation for Aggressive and Disruptive Behavior", *World Neurosurgery*, 80: S29.e11-S29.e14.

Franzini A., Marras C., Ferroli P., et al. 2005, "Stimulation of the Posterior Hypothalamus for Medically Intractable Impulsive and Violent Behavior", *Stereotactic and Functional Neurosurgery*, 83: 63–66.

Fuselli S. 2020, "Mental integrity protection in the neuro-era. Legal challenges and philosophical background", *BioLaw Journal – Rivista di BioDiritto*: 413–29.

Fuselli S. 2016, "Dalla neurobiologia al diritto: natura facit saltus", in S. Fuselli (ed) 2016, *Neurodiritto. Prospettive epistemologiche, antropologiche e biogiuridiche*, Milano: Mimesis: 105–34.

Gilbert F. 2015, "A Threat to Autonomy? The Intrusion of Predictive Brain Implants", *AJOB Neuroscience*, 6: 4–11.

Greely H.T. 2008, "Neuroscience and criminal justice: Not responsibility but treatment", *University of Kansas Law Review*, 56:1103–38.

Greely H.T. 2009, "Law and the Revolution in Neuroscience: An Early Look at the Field", *Akron Law Review*, 43: art. 2.

Greene J.D. and J. Cohen 2004, "For the law, neuroscience changes nothing and everything", *Philosophical Transactions of the Royal Society of London*, 359: 1775–1785.

Guger, C., Allison, B. Z., and Mrachacz-Kersting N. (2019), "Brain-computer interface research: A state-of-the-art summary 7", *Brain-Computer Interface Research* (pp. 1-9). Springer, Cham.

Hübner D. and L. White 2016, "Neurosurgery for psychopaths? An ethical analysis", *AJOB Neuroscience*, 7: 140–49.

Ienca M. and R. Andorno 2017, "Towards new human rights in the age of neuroscience and neurotechnology", *Life Sciences, Society and Policy*, 13: 5.

Kirchmair L. 2019, "Objections to Coercive Neurocorrectives for Criminal Offenders – Why Offenders' Human Rights Should Fundamentally Come First", *Criminal Justice Ethics*, 38:19–40.

Lavazza A. 2018a, "If Criminal Intentions Are Nonvoluntary, Mandatory Neurointerventions Might Be Permissible", *AJOB Neuroscience*, 9: 154–156.

Lavazza A. 2018b, "Freedom of Thought and Mental Integrity: The Moral Requirements for Any Neural Prosthesis", *Frontiers in neuroscience*, 12: 82–82.

Levy N. 2007, *Neuroethics: Challenges for the 21st Century*, Cambridge: Cambridge University Press.

Libet B. 2004, *Mind Time. The Temporal Factor in Consciousness*, Cambridge (MA) – London: Harvard University Press.

Mackenzie R. 2016, "Deep Brain Stimulation for Psychopaths–A No Brainer", *AJOB Neuroscience*, 7: 137–139.

Matravers M. 2018, "The Importance of Context in Thinking About Crime-Preventing Neurointerventions", in D. Birks and T. Douglas (eds) 2018, *Treatment for Crime. Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford: Oxford University Press: 71–93.

McMahan J. 2018, "Moral Liability to 'Crime-Preventing Neurointervention'", in D. Birks and T. Douglas (eds) 2018, *Treatment for Crime. Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford: Oxford University Press: 117–23.

McMillan J. 2018, "Containing Violence and Controlling Desire", in D. Birks and T. Douglas (eds) 2018, *Treatment for Crime. Philosophical Essays on Neurointerventions in Criminal Justice*, Oxford: Oxford University Press: 224–34.

Morse S (2011b) Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty. Law, *Innovation and Technology*, 3: 209–228. https://doi.org/10.5235/175799611798204932.

Morse S., 2011b, "Lost in Translation?: An Essay on Law and Neuroscience", *Law and Neuroscience Current Legal Issues*, 2010, 13: 529–562.

Nunziante A.M. 2016, "La naturalizzazione del pensiero come tesi metafilosofica", in S. Fuselli (ed) 2016, *Neurodiritto. Prospettive epistemologiche, antropologiche e biogiuridiche*, Milano-Udine: Mimesis: 17–54.

Palk A.C. 2018, "Mandatory Neurointerventions Could Enhance the Mental Integrity of Certain Criminal Offenders", *AJOB Neuroscience*, 9: 150–52.

Pardo M.S. and D. Patterson 2013, *Minds, Brains, and Law. The Conceptual Foundations of Law and Neuroscience*, Oxford: Oxford University Press.

Petersen T.S. 2018, "Should neurotechnological treatments offered to offenders always be in their best interests?", *Journal of Medical Ethics*, 44: 32.

Petersen T.S. and K. Kragh 2017, "Should violent offenders be forced to undergo neurotechnological treatment? A critical discussion of the 'freedom of thought' objection", *Journal of Medical Ethics*, 43: 30.

Pugh J. and T. Douglas 2016, "Justifications for Non-Consensual Medical Intervention: From Infectious Disease Control to Criminal Rehabilitation", *Criminal Justice Ethics*, 35: 205–229.

Reti I.M. and Chang A.D. 2015, "Introduction to Brain Stimulation", *I.M. Reti, Brain Stimulation Methodologies and Interventions*, Hoboken (NJ): Blackwell: 1-12.

Rousseau J.J. 2002, The Social Contract, in S. Dunn (ed) 2002, *The Social Contract and The First and Second Discourses*, New Haven and London: Yale University Press: 149–254.

Ryberg J. 2015, "Predictive Brain Devices, Therapeutic Activation Systems, and Aggression", *AJOB Neuroscience*, 6: 36–38.

Ryberg J. 2016, "Deep Brain Stimulation, Psychopaths, and Punishment", *AJOB Neuroscience*, 7: 168–169.

Ryberg J. and T.S. Petersen 2013, "Neurotechnological Behavioural Treatment of Criminal Offenders–A Comment on Bomann-Larsen", *Neuroethics*, 6: 79–83.

Sententia W. 2004, "Neuroethical Considerations: Cognitive Liberty and Converging Technologies for Improving Human Cognition", *Annals of the New York Academy of Sciences*, 1013: 221–228.

Serb A., Corna A., George R., et al. 2020, "Memristive synapses connect brain and silicon spiking neurons", *Scientific Reports*, 10: 2590.

Shaw E. 2014, "Direct Brain Interventions and Responsibility Enhancement", *Criminal Law and Philosophy*, 8: 1–20.

Shaw E. 2018, "Retributivism and the Moral Enhancement of Criminals Through Brain Interventions", *Royal Institute of Philosophy Supplement*, 83: 251–270.

Shaw E. 2015, "The Use of Brain Interventions in Offender Rehabilitation Programs: Should It Be Mandatory,Voluntary, or Prohibited?", in J. Clausen and N. Levy (eds) 2015, *Handbook of Neuroethics*, Dordrecht: Springer Netherlands: 1381–1398.

Shniderman A.B. and L.B. Solberg 2018, "Mandatory Neurointervention: A Lesser Evil Than Incarceration?", *AJOB Neuroscience*, 9: 148–49.

Sommaggio P. 2016, "Neurocivilizzazione o libertà cognitiva?", in S. Fuselli (ed) 2016, *Neurodiritto. Prospettive epistemologiche, antropologiche e biogiuridiche*, Milano-Udine: Mimesis: 135–54.

Sommaggio P., Mazzocca M., Gerola A. and F. Ferro 2017, "Cognitive liberty. A first step towards a human neuro-rights declaration", *BioLaw Journal – Rivista di BioDiritto*, 11: 27-45.

Walsh E., Kühn S., Brass M., et al. 2010, "EEG activations during intentional inhibition of voluntary action: An electrophysiological correlate of self-control?", *Neuropsychologia*, 48: 619–626.

Wiseman H. 2016, *The myth of the moral brain: The limits of moral enhancement*, Cambridge and London: MIT Press.