T

Neurolaw and Punishment: a Naturalistic and Humanitarian View, and its Overlooked Perils

Andrea Lavazza

1. *Neurolaw as a naturalization of law*

Neurolaw is the approach that attempts to apply recent progress in neuroscience to the classic conceptions of law, often with the aim of pushing legal institutions (especially in criminal law) to be more in line with scientific knowledge[1]. This is essentially a process of naturalization *a' la* Quine applied to an area – law – that so far has been largely unaffected by naturalization. This also applies to punishment, its aims, its methods of implementation and its justification.

Two kinds of issues arise when applying neuroscientific findings to the law[2]. The first, called internal, are already being tackled by present institutions (for example, cases of imputability) and do not involve any major modifications, but only partial adjustments in some cases. A classic example is that of the legal age of majority, which can vary from system to system, and from country to country. The conventionalistic element is obviously predominant in the decision to place the age of legal responsibility at 18 rather than 16 or 21, but this choice has always been also linked to the psychological knowledge available at the time. Today, however, we know that the maturation of the prefrontal cortical areas of the brain, critical for controlling behavior and modulation of instinctive-impulse response, continues throughout adolescence and part of youth, until at least

[1] M.S. Pardo-D. Patterson, *Minds, Brains, and Law: The Conceptual Foundations of Law and Neuroscience*, Oxford University Press, New York 2013; D. Patterson-M.S. Pardo (eds.), *Philosophical Foundations of Law and Neuroscience*, Oxford University Press, New York 2016; A. Lavazza-L. Sammicheli, *Il delitto del cervello. La mente tra scienza e diritto*, Codice, Torino 2012.

[2] B.N. Waller, *Against Moral Responsibility*, MIT Press, Cambridge (MA) 2011.

age 20-22. This may have consequences for the decision whether or not to punish a young person who has committed certain types of crimes. It is no coincidence that the US Supreme Court, when deciding on the constitutionality of the death penalty for juveniles (Roper v. Simmons, 2005), also heard the opinions of neuroscientists. The decision to declare the death penalty for juveniles unconstitutional was not explicitly justified with neuroscientific findings, but many observers have expressed the belief that clinical data have had a significant role in it[3].

External issues, instead, are those involving the so-called *ius condendum*: the rewriting or radical reformulation of the main legal institutions based on the evidence provided by science, according to which such institutions and their underlying principles are no longer responsive to the known facts. Punishment belongs to this second category.

2. *The problem of free will*

A relevant line of naturalization of criminal law relies on the developments in neuroscience so as to try to prove that (if not always, at least most times) our actions are not free according to the classic definition of freedom – where the agent is capable of knowingly, voluntarily and *consciously* undertaking a course of action by choosing between alternatives. On the contrary, it is posited that our actions feature a high degree of determinism or at least of unconsciously undertaken courses of action, so that criminal conduct is regarded as deriving from the genetic asset of the subject, partly conjugated with an unfavourable environment. Other lines of research highlight that the structure and functioning of the brain strongly shape the subject's character traits (empathy in the first place) and can therefore direct or influence the behaviour of the individual in question[4].

More precisely, scepticism about free will is due to three main elements[5]. The first is the classical objection to freedom: determinism,

[3] D.L. Faigman-O.D. Jones-A.D. Wagner-M.E. Raichle, *Neuroscientists in Court*, in «Nature Reviews Neuroscience», 14 (2013), n. 10, pp. 730-736.

[4] A.R. Cashmore, *The Lucretian Swerve: The Biological Basis of Human Behavior and the Criminal Justice System*, in «Proceedings of the National Academy of Sciences», 107 (2010), n. 10, pp. 4499-4504; P.S. Churchland, *Braintrust: What Neuroscience Tells Us about Morality*, Princeton UP, Princeton (NJ) 2011.

[5] G.D. Caruso, *Introduction: Exploring the Illusion of Free Will and Moral Responsibility*, in G.D. Caruso (ed.), *Exploring the Illusion of Free Will and Moral Responsibility*, Lexington Books, Lanham (MD) 2013, pp. 1-16.

declined in several forms. The second, supported by the majority of philosophers of the mind, is the impossibility of mental causation, which is a condition for agency causation, a fundamental part of libertarian positions. The third is given by recent findings of cognitive science, indicating a progressive breakdown of the conscious self (some experiments seem to completely disconnect the latter from so-called "free" choices). In this regard, Nahmias underlines that this third strand is specifically interested in the progress of empirical psychology and cognitive neuroscience. In particular, he considers the first two strands as related to the form of mental causation, while the last is a thesis on the *content* of mental causation[6].

In cognitive science (including neuroscience) there is an ongoing process that is in line with this trend I have just described: the process of "deconstruction" of the conscious and rational unitary self – the subject of free will. Here one can distinguish two subsets. One concerns the beginning of the action: conscious intentions are preceded by subconscious cerebral processes[7]; the other concerns the conscious control of behavior, stating that consciousness is unaware of the automatic processes at work and the true reasons for our conduct[8]. The point is essentially that, under a more thorough empirical examination, more often than we would think, cognitive processing appears to be the result of subpersonal processes of which we are unaware.

These are automatic processes, triggered by the environment or the situation, bound to a repertoire that is partly innate and partly due to experience and education; such processes causes bodily responses due both to the tendency to homeostasis and to the search for what is functional to our survival and physical and mental well-being[9]. There are many examples of this decomposition of the self into cerebral modules that elaborate infor-

[6] E. Nahmias, *Is Free Will an Illusion? Confronting Challenges from the Modern Mind Sciences*, in W. Sinnott-Armstrong (ed.), *Moral Psychology. Vol. 4 Free Will and Moral Responsibility*, MIT Press, Cambridge (MA) 2014, pp. 1-25.

[7] In this respect, think of the very famous studies by Benjamin Libet: cf. B. Libet-C.A. Gleason-E.W. Wright-D.K. Pearl, *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act*, in «Brain», 106 (1983), n. 3, pp. 623-642; B. Libet, *Mind Time: The Temporal Factor in Consciousness*, Harvard UP, Cambridge (MA) 2004.

[8] Peter Carruthers is one of the most consistent supporters of this line of thought; P. Carruthers, *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford UP, New York 2011; Id., *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*, Oxford UP, New York 2015.

[9] See, for example, J.M. Doris, *Talking to Our Selves*, Oxford UP, New York 2015.

mation autonomously and subconsciously, which then emerge as a single apparent stream of consciousness. One case is that of language, where all the processes that lead us to say the words we speak are completely opaque to our consciousness[10].

Nevertheless, there is still wide consensus that neither recent experimental research through EEG and brain imaging, nor evidence coming from empirical psychology are enough to conclusively state that human beings have no free will[11]. Recent interpretations of the data collected by Libet even seem to bring back brain mechanisms of free will similar to our intuitive conception of it[12], which would also allow for a better understanding of it in terms of legal applications[13].

3. *Free will, law, and punishment*

One of the most discussed arguments regarding the notion of free will as an illusion and its consequences on the law is the one developed by Greene and Cohen[14]. According to their argument, a truly scientific description of the human being is incompatible with the attribution of *pure desert* in relation to the decisions made by the subject, on the basis of which the legitimacy (and effectiveness) of legal sanctions is determined. The proponents of this view maintain that one cannot but follow the logical

[10] T. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious*, MIT Press, Cambridge (MA) 2002.

[11] A.R. Mele, *Effective Intentions: The Power of Conscious Will*, Oxford UP, New York 2009; Id., *Free: Why Science Hasn't Disproved Free Will*, Oxford UP, New York 2014.

[12] A. Schurger-J.D. Sitt-S. Dehaene, *An Accumulator Model for Spontaneous Neural Activity Prior to Self-initiated Movement*, in «Proceedings of the National Academy of Sciences», 109 (2012), n. 42, pp. E2904-E2913; A. Schurger-M. Mylopoulos-D. Rosenthal, *Neural Antecedents of Spontaneous Voluntary Movement: a New Perspective*, in «Trends in Cognitive Sciences», 20 (2016), n. 2, pp. 77-79.

[13] A. Lavazza-S. Inglese, *Operationalizing and Measuring (a Kind of) Free Will (and Responsibility). Towards a New Framework for Psychology, Ethics and Law*, in «Rivista Internazionale di Filosofia e Psicologia», 6 (2015), n. 1, pp. 37-55; A. Lavazza, *Free Will and Neuroscience: From Explaining Freedom Away to New Ways of Operationalizing and Measuring It*, in «Frontiers in Human Neuroscience» 10 (2016) art. 262.

[14] J. Greene-J. Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, in «Philosophical Transactions of the Royal Society of London B: Biological Sciences», 359 (2004), n. 1451, pp. 1775-1785. But see also R. Sapolsky, *The Frontal Cortex and the Criminal Justice System*, in S. Zeki-O.R. Goodenough (eds.), *Law and the Brain*, Oxford University Press, Oxford 2004, pp. 227-243; and S. Harris, *Free Will*, Simon and Schuster, New York 2012.

sequence deriving from the experimental data, which for them leads to the unavoidable pragmatic conclusion of choosing a consequentialistic kind of law and punishment.

According to Greene and Cohen, although Western penal systems claim to be compatibilist regarding free will, they actually seem to presuppose a libertarian perspective. But this view is now being threatened by the findings of neuroscience, which refers to a form of brain-related determinism. This kind of scientific data is in contrast with the widespread commonsense view of justice as well as with the retributivist conception of the law. Knowledge of the functioning of the brain points in the direction of denying the concept of free will in those who commit crimes, therefore leading to consequentialism: a view that – according to its promoters – is more in line with the scientific description of the human being.

Interestingly, the consequentialist perspective that relies on the idea of free will as an illusion also disrupts the limitations to the most undesirable consequences of the classical utilitarian perspective on punishment, which did not have arguments, for example, to exclude the use of scapegoats in some extreme cases. Among others, this "preventive" argument is supported by Hart[15]. Let's have a look at its logical path as it was retraced by De Caro and Marraffa[16].

In retributivism it is possible to identify two components, one called positive (all the guilty deserve to be punished with the required severity) and one called negative (only the guilty deserve to be punished, with no excessive severity). The second element prohibits to punish those who do not deserve it, and also has a preventive element against inhuman and disproportionate sentences. One could claim that the two components are logically independent, so that only one of them can be adopted. That is what Hart does, justifying punishment in purely utilitarian terms and using the negative component of retributivism as the "limit" to respect when attributing punishment, so as to avoid cases of blatant injustice. In other words, one can never punish an innocent, or someone who is causally but not morally responsible for a bad deed (say, because they are unfit to plead). This also holds when the punishment would have beneficial consequences for society as a whole.

[15] H.L.A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law*, Oxford UP, Oxford 1968.

[16] M. De Caro-M. Marraffa, *Mente e morale. Una piccola introduzione*, Luiss UP, Rome 2016, pp. 98-102.

For Hart, however, the justification of the sentence is not based on the merits of the offender (that is, when one punishes someone, it's not because there is an assumed balance of justice to be restored or because the act of the offender has to be punished as such). Hart refers to consequentialist model, according to which the only justification for punishment is that it is socially useful. Punishments thus serve: to create a deterrent so that, under the threat of punishment, people refrain from committing crimes; to make sure that dangerous people (because they have committed crimes) are put in a position not to further harm society; and to make criminals fit for social life through the execution of the sentence.

However, if you give up the chain that from the possibility to do otherwise – the primary condition of free will – leads to the idea of moral responsibility (understood as more than a contribution to the physical causal process of an event) then the negative retribution clause is no longer relevant. Therefore, there is no reason why classical utilitarianism shouldn't reappear at its purest, justifying the punishment of an innocent if it benefits the majority, as there are no principles against it other than merely conventional ones. A human being unable to act otherwise might be attributed some other form of dignity, but when it comes to punishment it is hard to see how the notion of "responsibility" can be replaced[17].

Indeed, as Greene and Cohen put it, given determinism in its various forms, consequentialism works in every case[18]. This is because this concept does not pose the problem of someone being truly innocent or guilty in some ultimate sense that depends on the freedom of action, but only addresses the issue of the likely effects of the punishment (although there is the problem of absolute determinism that does not seem to allow for the deterrent effect). The retributivist approach, on the contrary, seems to require the idea of free will, namely the ability to do otherwise, which is the classic condition for responsibility. If every action is the result of brain mechanisms outside of the possible conscious control of the subject, mechanisms visible through techniques capable of seeing in the "transparent bottleneck" of our nervous system, then it makes no sense to blame choices and actions on the subject who makes them. In a way, according to

[17] G. Sartori-A. Lavazza-L. Sammicheli, *Cervello, diritto e giustizia*, in A. Lavazza-G. Sartori (eds.), *Neuroetica. Scienze del cervello, filosofia e libero arbitrio*, il Mulino, Bologna 2011, pp. 135-163; A. Lavazza- L. Sammicheli, *Se non siamo liberi, possiamo essere puniti?*, in M. De Caro-A. Lavazza-G. Sartori (eds.), *Siamo davvero liberi?*, Codice, Torino 2010, pp. 147-166.

[18] J. Greene-J. Cohen, *op. cit*.

the supporters of free will as an illusion, those choices and actions are the result of an automatic response to social and environmental stimuli or internal impulses oriented to the preservation and to the physical and psychological wellbeing of the individual.

4. *The consequentialist view and its perils*

The outcome of this view is that punishment should be detached from any retributivist justification: it should not be afflictive in its main purpose, because this goes against the humanitarian principles of not harming our fellows without reason. In fact, the mere enlargement of the category of non-liability due to the discovery that many of those who commit violent crimes have serious brain abnormalities would lead to suspend or eliminate classical punishment in favour of other protective measures, such that would not be afflictive and would not have the sole purpose of punishing evil with more evil[19]. Along the same lines, the goal of rehabilitation of classical punishment would also cease to exist for people who are "mad and not bad", so to speak.

Derk Pereboom is perhaps the most important supporter of this thesis[20]. According to Pereboom, who is a hard incompatibilist[21], "living without free will" and, therefore, without responsibility, does not affect our ideas of morality, meaning and value of existence; so it's not something that produces the upheavals feared by defenders of free will. His "Spinozan" idea is that the main effects would be the end of a retributivist penal system (based on what the individual has done before) and the abolition of excessively severe punishments including, of course, the death penalty.

However, the adoption of a consequentialist perspective, inspired by crime prevention and social rehabilitation, would not exclude measures such as preventive detention. Pereboom considers the latter an instrument of social protection morally comparable to quarantine for people with highly contagious diseases. Just as the sick are not responsible for their condition,

[19] Cf. K. Kiehl, *The Psychopath Whisperer: Inside the Minds of Those Without a Conscience*, Oneworld, New York 2015.

[20] D. Pereboom, *Living without Free Will*, Cambridge UP, Cambridge 2001; Id., *Free Will, Agency, and Meaning in Life*, Oxford UP, New York 2014.

[21] Hard incompatibilists state that both determinism and indeterminism argue to the detriment of freedom, since in both cases the behavior of the subjects is caused by factors that are beyond their control.

and yet can be isolated for as long as necessary, guaranteeing them the best care and the highest personal dignity, so deemed dangerous subjects can be given the opportunity to do no harm without further afflictions, indeed, favoring their recovery. Pereboom's view is marked by the overcoming of "moral rage", which, for him, damages both individual wellbeing and interpersonal relationships due to the persistent tendency to blame and reprimand people, with the consequent creation of moral "debit" and "credit" able to poison one's life.

In a fully developed neurolaw, then, punishment would never be a substitute for a sort of social revenge, but rather the most humane tool available to control dangerous subjects and protect society, based on the medical and neuroscientific knowledge available. For sex offenders, for instance, it could be possible to act drastically with drugs that lower the hormone levels relevant to the behaviour in question (chemical castration); for impulsive and violent subject, drugs that control one's mood would be appropriate. In other cases, brain pacemakers, in the form of brain stimulators, may act by reducing certain compulsive behaviours (such as taking drugs that lead to other crimes), and so forth.

An approach of this kind would be welcomed both by society and by the very individuals subjected to it, because it would be selective and would not completely deprive them of their physical freedom (or of their life, where death penalty is in force). Nevertheless, this approach has the characteristic of potentially slipping into (1) the invasive violation of privacy and bodily integrity (the right that protects against intentional interference with one's body) on the basis of available technology; (2) preventive treatment or detention; (3) treatment or detention without a specific goal. This could happen if punishment were no longer anchored to the classical mechanism of personal responsibility in the retributive sense, for which one is punished for what one has done in accordance with a law that pre-established the punishment according to the seriousness of the crime as such. But the consequentialist perspective tends to radically break away from that model.

Let's see in more detail the three forms of punishment implied by a consequentialism that denies any retributivist element. These are forms of punishment that conflict with strong moral intuitions and violate ethical principles that seem to be a shared heritage for the defense of the individual and her autonomy. Preventive treatment or detention could be put in place for those that, on the basis of neuroscientific markers and other behavioral data, are expected to have high chances of committing violent crimes. To this end, mass screening would become mandatory from birth,

and this could open the door to discrimination and strong personal autonomy limitations. If the subject shows predictive markers of serious antisocial behavior, according to the consequentialist perspective, he should be put in a condition to do no harm, as the focus is entirely on the protection of the community at the expense of the single potential criminal. The subject is indeed denied many rights in the implicit assumption that he is a "sick" person, who should be treated as for her well-being but deprived of physical freedom and self-determination, in order to protect society.

From chemical castration to genetic engineering, all systems of care to improve deviant behavior would become lawful. Furthermore, such care or indefinite periods of detention would not have a clear goal, since at least for some individuals the point would be to prevent the general threat to society that they could potentially represent – contrary to the retributivist system, there would be no need to wait for the threat to be actualized. Ultimately, the availability and justifiability of this new kind of punishment might lead to apply it to *all* those individuals who have been identified as highly likely to commit certain crimes: one might want to coercively treat with drugs both an exhibitionist and a rapist. Secondly, whenever a technique promised to be efficient in detecting or preventing criminal conduct, it would be justified to introduce it and enforce it on potentially interested parties. Thirdly, in some cases it is unclear how to assess the decreased dangerousness of subjects under coercive treatment, so that the treatment could be extended indefinitely.

It is useful to recall here the position expressed by Thomas Douglas. He has persuasively argued for criminal rehabilitation through medical interventions (such as medications that replace the drug of addiction for drug-addicted offenders, and injections of testosterone-lowering drugs for sex offenders) claiming that committing a crime can render one morally liable to certain forms of medical intervention[22]. Douglas challenges the shared assumption that medical interventions may only permissibly be administered to criminal offenders with their consent. The argument starts from the fact that it is commonly accepted that the State can impose a punishment without consent to those who commit crimes, typically a period of detention. For Douglas, if one accepts that offenders are morally liable, imposing limitations on freedom of movement and association (with all

[22] T. Douglas, *Criminal Rehabilitation Through Medical Intervention: Moral Liability and the Right to Bodily Integrity*, in «The Journal of Ethics», 18 (2014), n. 2, pp. 101-122.

that this implies) does not produce more harm than a violation of bodily integrity, when it is oriented to the rehabilitation of the offender.

Nevertheless, there is still the problem of direct brain interventions: contrary to the lowering of testosterone to temporarily reduce sexual desire, such interventions interfere with the very basis of agency and the self. As Jared Craig rightly put it, there is a "more fundamental right to 'mental integrity'" that defends an alleged inner sphere of liberty and protects critical capacities necessary for the exercise of autonomous human agency – without which a vast majority of moral rights could not exist[23]. In this sense, the State should not be entitled to administer direct brain interventions to criminal offenders without a valid consent.

Here's an example of consequentialist scenario sketched by Adrian Raine:

> Under LOMBROSO [program – Legal Offensive on Murder: Brain Research Operation for the Screening of the Offenders], all males in society aged eighteen and over have to register at their local hospital for a quick brain scan and DNA testing. [...] The result is not a perfect prediction, but it is pretty darn good [...] Those classified as LP-S (Lombroso Positive-Sex) have an 82 percent chance of committing either rape or pedophilic offenses. [...] The program works like this: those who test positive – the LP-S – are held in indefinite detention[24].

But the program can be expanded.

> Poor parenting has undeniably been linked to later violence. Genetic studies documented not just that antisocial parents transmit their bad genes to their children, but that negative social experience of having a bad parent is also a causal factor for antisocial behaviour. [...] Cars can be killers, and so you need a licence before you can drive. Kids can be killers too. So the logic goes that you should also have a licence before you can have a child[25].

Then, even something with a scientific justification and a related humanitarian goal could dangerously turn into an instrument of tyranny and discrimination, because the scientific knowledge in this area is not yet well

[23] J.N. Craig, *Incarceration, Direct Brain Intervention, and the Right to Mental Integrity. A Reply to Thomas Douglas*, in «Neuroethics», 9 (2016), n. 2, pp. 107-118; cf. also J.C. Bublitz-R. Merkel, *Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-determination*, in «Criminal Law and Philosophy», 8 (2014), n. 1, pp. 51-77.

[24] A. Raine, *The Anatomy of Violence: The Biological Roots of Crime*, Pantheon, New York 2013, pp. 342-343.

[25] Raine, *op. cit.*, p. 349.

substantiated. Also, the laws would be adopted by political decision-makers without all the necessary technical knowledge, and judgments would be made not only by experts but also by judges guided by considerations other than the simple medical history of the defendant. Indeed, the very fact that the new type of punishment has an allegedly more humane character would end up lowering the public attention to potential miscarriages of justice.

5. *The Strawsonian view*

Another set of (more philosophical) considerations appeals to a perspective proposed by Peter Strawson, according to whom the naturalistic-consequentialistic approach treats the human beings subject to the new type of punishment as broken machines rather than as agents to be respected and considered worthy of dignity[26]. In *Freedom and Resentment*, Strawson considers "the non-detached attitudes and reactions of people directly involved in transactions with each other", or else "the attitudes and reactions of offended parties and beneficiaries; of such things as gratitude, resentment, forgiveness, love, and hurt feelings"[27]. In our interactions with our fellow human beings, we all have reactive attitudes and feelings, which we ourselves are subject to. They have an extraordinary importance for us and depend on what we think about the feelings and attitudes of others.

Resentment towards people who deliberately harm us is not philosophically problematic; however, there are two factors that could affect that feeling if those who harmed us did so under particular circumstances. The first one is related to unintentionality: "He didn't mean to", "He hadn't realized", "He was pushed". In such cases we might curb our resentment but still feel that it's appropriate to have a reactive response. The second one is related to cases when the person responsible "wasn't himself", "has been under very great strain recently", or even "is a hopeless schizophrenic", "his mind has been systematically perverted". For Strawson, such cases lead us to restrain from having our normal reaction towards the agent. Hence a contraposition between participation/involvement in a human relationship and what could be called an "objective attitude" towards other humans.

[26] P.F. Strawson, *Freedom and Resentment*, in «Proceedings of the British Academy», 48 (1962), pp. 1-25.

[27] *Ivi*, pp. 82-83. I am here taking up an exposition found in A. Lavazza-L. Sammicheli, *op. cit.*, cap. 8.

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained [...] The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships[28].

That is, if one adopts the objective attitude towards someone, feelings such as resentment, gratitude, forgiveness, anger or romantic love cannot arise. One can talk with that person, but not argue. In other words, we would say, one does not perceive them as able to respond to reason. If one accepts determinism (incompatibilism), then, should one give up the reactive attitudes? The answer is that this would be impossible, because of the very way we are made: the involvement with which human beings participate in common interpersonal relationships is too intense and runs too deep to seriously believe that a general theoretical conviction might genuinely change our world – including interpersonal relationships as we normally understand them[29].

But one may ask: what would be the rational choice, if freedom were really illusory? According to Strawson, firstly, we are *naturally* led to reactive attitudes, we cannot choose whether or not to adopt them in the way that we can, for example, accommodate or not some preferences; secondly, and most importantly, even if we had a choice, the rational option would be to evaluate gains and losses for human life, considering what enriches or impoverishes it; the truth or falsity of a general thesis related to determinism would not have any relation to the rationality of this choice. Personal reactive attitudes are based on an expectation and a need: that human beings around us show a certain degree of goodwill or regard towards us; or, at least, that they show no active manifestation of malice or indifference. It follows that it is simply *useless* to ask whether or not it would be rational for us to actualize something that by virtue of our own nature we cannot

[28] Strawson, *op. cit*., pp. 89-90.

[29] B. Vilhauer (*The People Problem*, in G.D. Caruso (ed.), *Exploring the Illusion of Free Will and Moral Responsibility*, Lexington Books, Lanham (MD) 2013, pp. 141-160) believes that one can overcome Strawson's argument on the depersonalization of human beings by referring to the Kantian principle that prescribes to treat all our fellow beings always as ends and never as means. This principle can be declined without the use of reactive attitudes and attributions of freedom and moral responsibility.

(be able to) do. The general network of personal reactive attitudes was in fact created along with human society and, considered as a whole, does not need any external *rational* justification.

As is obvious, even regardless of the facts in favor of the deterministic thesis, common sense and scientific optimism given by the illusionary character of free will clash with the fact that one wants to choose on the basis of practical consequences of her decision. Finally, from their point of view, the skeptics who are optimistic on free will are those who confirm that it is possible to have a humanitarian "objective attitude" oriented to the welfare of others. But from the point of view of the "optimistic skeptics" this is contradictory to what previously stated on consequentialism, unless one introduces the purpose of respecting certain values that are themselves disjointed from consequentialism itself.

6. *A defence of moderate retributivism*

Given the perils coming from a purely consequentialistic perspective denying free will and responsibility, I briefly propose three arguments to defend a moderate conventionalist thesis on the classical responsibility of retributivist law. I think it is appropriate to maintain (by stipulation, according to the liberal-democratic processes that form and gradually change the legal system) a system that recognizes – mostly and with encoded exception types – freedom, rationality and the ability to answer for one's actions. Such a system also includes punishments directly related to the voluntary transgression of the norms, although also aimed at the recovery of the offender and the protection of society.

The first argument is related to naturalism, the very frame in which refoundational prospects are inscribed. The evolutionary processes of the species, driven by selection and adaptation, have endowed us with very strong intuitions – generally retributive – that cause people to be ready to bear a personal cost, with no other gain than the restoration of a sense of justice, to punish offenders who deserve it[30]. It does not seem easy to reverse this intuition with a rationalist argument, especially as one looks at the same time to found morality on the feelings of "natural sympathy" that

[30] E. Fehr-S. Gächter, *Altruistic Punishment in Humans*, in «Nature», 415 (2002), n. 6868, pp. 137-140; E. Fehr-U. Fischbacher, *Third Party Punishment and Social Norms*, in «Evolution and Human Behavior», 25 (2004), n. 2, pp. 63-87; cf. also A. Lavazza-L. Sammicheli, *op. cit.*, cap. 7.

are probably the result of evolution[31], thus replacing ethical systems *a la* Kant[32].

The second (indirect) argument is related to experiments aimed to falsify common sense and naive psychology, which provide the basis to naturalistically falsify retributivism and give arguments in favor of consequentialism denying the intuitions of freedom and responsibility. The impression is that these experiments are, so to speak, "below threshold" with respect to relevant social macro-interactions for the dynamics of allocation of responsibility and the functioning of relevant interpersonal relations. In this sense, the relationship between the description of the subpersonal mental mechanisms and intentionalist psychology could be seen in analogy with what happens in physics between relativistic mechanics and classical mechanics. Relativistic mechanics is certainly more appropriate to the current knowledge and allows for a "true" and finer description of reality, but the most intuitive and usual description offered by Newton's classical mechanics is perfectly adequate for most of the macroscopic applications that usually concern us. As for the description of human beings there is also a subjective element, which seems to prefer – for now, but for many reasons – the use in certain areas of folk psychology. Also, one could say that what allows empirical psychology to describe the disunity of the subject and the alleged behavioural automaticness is a "quantification" that covers a narrow area of our spectrum of social action. The significant interactions subject to the law might fall within the macroscopic range of relevant values in which behaviour tends to be free, aware and rational – that is, coherent with the assumptions of retributivism.

[31] S. Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford UP, New York 2004; E. Lecaldano, *Prima lezione di filosofia morale*, Laterza, Roma-Bari 2011; Id., *Simpatia*, Cortina, Milano 2013.

[32] Robert Nozick presents a theoretically refined version of common sense: "In terms of the connection with value effected by punishment we can understand some of the metaphors that stud retributivist talk. Wrong puts thing out of joint in that acts and persons are unlinked with correct values; this is the disharmony introduced by wrongdoing. Punishment does not wipe out the wrong, the past is not changed, but the disconnection with the value is repaired (though in a second best way); nonlinkage is eradicated. Also, the penalty wipes out or attenuates the wrongdoer's link with incorrect values, so that he now regrets having followed them or at least is less pleased that he did" (R. Nozick, *Philosophical Explanations*, Harvard UP, Cambridge (MA) 1981, p. 379). Retributivism, just as the consequentialism evidenced by Hart, needs external principles to define its scope. Nozick himself recognizes this implicitly when he asks why we should not always relate to the value, even for those who do not commit crimes. The answer is that in that case what prevails is the individual's right to be left alone.

The third argument relies on a distinction that has been used to show how psychopaths could be considered exempt from the law (also ethically) by virtue of the fact that they fail to grasp the strength of genuinely moral prescriptions, therefore lacking the ability to understand the scope and consequences of their actions[33]. This concerns the partition between conventional norms (you have to sit straight at school) and moral norms (you mustn't pull your classmate's hair), which also small children are able to grasp. Now, if this distinction is based on some foundation, related to a specific ability of recognition, this seems to imply some form of moral realism. Not a realism that presupposes an autonomous existence of values that people can grasp with a special sense, but more likely a pre-reflective intuition shared by almost all human beings on the evaluation of a series of behaviors as a positive or negative (to do or not to do).

If these insights may serve as a point of reference and constitute a reason for exemption for those who, due to a "natural" defect, do not have them, then they must have a "validity" that enables them to act as a reference for "moral" behaviors. Those without this ability cannot be held responsible; conversely, those who do, though, when not respecting these rules, having the ability to understand them and to respect them, are exposed to reprimand and punishment. One could say that being able to grasp moral norms does not in itself amount to being able to respect them. However, in the psychopath argument, it must obviously be so, otherwise it would not make sense to use it so as to separate her position from that of other individuals. If there wasn't at least someone able to grasp moral norms and respect them, it would not make sense to declare that others (psychopaths) are instead exempt from them. If no one has the ability to respect the rules, then all, without distinction, should be declared exempt.

7. *Conclusion*

In a Kantian-Hegelian sense, punishment amounts to recognizing the freedom of action of the other, who more or less voluntarily has broken the law. The "bad person" is like us and can be "rehabilitated" with equal treatment. But quarantine, with the physical removal of the "bad" from

[33] E. Turiel, *The Development of Social Knowledge: Morality and Convention*, Cambridge UP, Cambridge 1983; R.J.R. Blair-D.G.V. Mitchell-K. Blair, *The Psychopath: Emotion and the Brain*, Blackwell, Oxford 2005.

society, is also a metaphor for the expulsion of the "sick". A "sick person" (socially non-integrated) that cannot be cured is a substantially different subject that can be legitimately treated as such.

In fact, Kant and Hegel have defended the retributivist principle, regardless of its roots in free will, as an instrument of protection of human dignity, which recognizes rational agency as constitutive of the person. In this view, the sign of autonomy – denied by the idea of criminals as sick, for whom the only option is extrinsic care – lies precisely in the ability of moral redemption through punishment. This does not mean that we should oppose the prospect of a neuroscientific punishment on a consequentialist basis, but rather that we should assess the risks and benefits of such an approach in the light of the full spectrum of punishments, their goals and their justifications. Mostly, we shouldn't fail to consider some principles that appear important for the dignity and autonomy of every human being as a subject endowed with intrinsic value.

Abstract

Neurolaw is the approach that attempts to apply recent progress in neuroscience to the classical conceptions of law, often with the aim of pushing legal institutions (especially in criminal law) to be more in line with scientific knowledge. It is essentially a process of naturalization of the law, which also applies to punishment, its aims, its methods of implementation and its justification.

A relevant line of naturalization of criminal law relies on developments in neuroscience so as to try to prove that (if not always, at least most times) our actions are not *free according to the classic definition of freedom – where the agent is capable of knowingly, voluntarily and* consciously *undertaking a course of action by choosing between alternatives. According to the proponents of this view, one cannot but follow the logical sequence deriving from the experimental data, which leads to the unavoidable pragmatic conclusion of choosing a consequentialistic kind of law and punishment.*

Consequentialist punishment is deemed to be more humane because it is not afflictive and is only targeted to protect society. But the fact that the charged person is regarded as more mad than bad, so to speak, turns her into a sort of "broken machine", with the risk of legitimizing preventive treatments or ones of indefinite duration. The objections to this approach are therefore related to the gaps of knowledge we still have, to the risks of "political"

abuse, and to the Strawsonian line of thought for which we cannot treat our fellow human beings as broken machines to be repaired, depriving them of their nature of free and rational agents (except in exceptional and rare cases). I suggest a more nuanced assessment of these possible developments and defend a moderate form of retributivism.

Keywords: free will; consequentialism; retributivism; naturalism; self; P. Strawson.

Andrea Lavazza
Centro Universitario Internazionale -Arezzo
lavazza67@gmail.com