Przemysław Zawadzki's doctoral thesis titled "The Ethics of Neuromodulation of the Self: Personal Identity, Authenticity, Autonomy and Moral Responsibility in the Light of Neurointerventions" presents itself as a work of considerable scholarly value, the result of a long course of research. Let me at the beginning of this my review on the doctoral thesis note a difference between the approach followed by Phd candidate Zawadzki compared to current academic practice in Italy, the country I am academically based in.

Highlighting the difference is also a way of introducing a mode that I think is useful in analyzing the thesis itself. In Italy the doctoral candidate, especially in the field of philosophy and the humanities in general, is required to write a real monograph, a work equivalent to a book of 150-200 pages on average. This is the development of the research project that should occupy the candidate admitted to the doctoral program for the entire period of study and work. The project to be developed is the one proposed by the candidate and on the basis of which he or she was selected for the doctoral course. Thereafter, the candidate is followed by their supervisors in writing the work, which must have, as mentioned, the structure of an academic volume, divided into chapters. In most cases, the doctoral thesis produced in this form is not published. A percentage that I could not quantify but could be as high as 15-20% gets the chance to be published, almost always in Italian by publishers who guarantee the book a small circulation. Often authors, after getting their Ph.D., edit and try to make their thesis more solid in order to publish part or all of it. Not all of them succeed. Their work was not in vain, even if it is not published, it still served the purpose of their maturation as scholars.

Przemysław Zawadzki's thesis appears "strange" in the eyes of an evaluator bound by the model I described above. But once its structure is understood it also appears worthy of admiration. It is a doctoral dissertation already published before it has even been assembled. It brings together nine scholarly articles published in peer-reviewed journals, almost all of them first-rate in their field, namely that of bioethics, neuroethics and applied ethics. This means that the author has already achieved that ability to interact with the international community of scholars in his field sufficient to have nine articles published in a rather short period of time, which is uncommon for a young scholar at the beginning of his academic career.

We know what the criteria are for publication in a scientific journal that has blind reviewing. In general, the article must support an original and coherent thesis, must contain good and sound arguments, must consider the relevant literature in its field, and must parry the most common possible objections to its main thesis. However, peer review is not always the most objective and reliable one can hope to have. I think it is not unfair to make this criticism of the review process in academic journals.

Often all of us, even experienced authors, come across unfounded or unfair criticism, scholars who do not know the subject matter well on the topic issue unquestionable judgments, and so on. Of course, we pay attention to these distortions when the judgment is negative and forces us to major revisions or even is such that it results in rejection. When it happens to us that a reviewer suggests "accept as is" we usually do not pose these problems and do not think that they may have been wrong in their assessment. This is normal; we have a positive bias toward our work.

Here I think the question must be asked whether the seal of quality and originality that publication in authoritative journals gives to the nine chapters of candidate Zawadzki's thesis is fully deserved. Experience tells me that when you publish nine papers in three years (2020-2022) in journals such as those in which candidate Zawadzki has published, the risk of a lucky chance is reduced to almost zero. I can reveal here, I don't think it's anything incorrect at this point, that I myself was the anonymous reviewer of one of the articles in the thesis. I will not say which one. I am a fairly active reviewer, I think I conduct a higher average number of reviews per year than my colleagues, and the topics of Zawadzki's papers are within the scope of my core competencies. In general, I am not a particularly strict reviewer, but I try to be honest and fair, taking into consideration the quality of the work and the journal in which it should be published, without being influenced by anything else.

When I received Zawadzki's manuscript the impression was very good. Of course, at the time I did not know who the author of the article was, but after publication I learned that it was the work of the candidate. The text was very well written and argued for, in excellent English, without any sloppiness. I considered suggesting some changes and additions in line with what I thought was a proposal to improve an article that was already of a sufficient standard. As is well known, articles published in peer-reviewed journals almost always end up coming out better than when they were first submitted.

Peer review should have above all the function of gatekeepers, that is, ensuring that only articles that possess scientific soundness are published, in other words, made public with a guarantee of that scientific soundness. This is how all disciplines progress (I will say more about the specifics of philosophy in this regard, an element that is also relevant in the evaluation of candidate Zawadzki's thesis). However, the review process also helps to correct small or large errors and inaccuracies, to refine arguments and form. This is the contribution that reviewers make to their colleagues and to the scientific community as a whole.

This methodological premise about peer review serves to say that the nine chapters that make up Zawadzki's thesis are the result of a collective process that contributed to their current form and final quality. At the same time, it means that the initial level of the manuscripts submitted to the journals that later published them was good

enough for the reviewers to give their positive opinion, albeit with some suggestions for improvement.

The part that deals specifically with philosophy, as mentioned earlier, has to do directly with the issues Zawadzki addresses. Generally, it is said that a reviewing in philosophy should be based on the soundness of the argument. According to this perspective, it is not up to the reviewer to judge the outcome of the argument, provided it is well developed. In other words, it matters how one defends a particular claim, not what the content of that claim is. But it is common experience that this is not the case. The evaluation often takes a personal dimension; the reviewer does not recognize the value of a contribution that he or she does not agree with and thus rejects the article because its conclusions do not fit with their philosophical perspective on the topic at hand.

This, paradoxically, is another element in favor of the quality and originality of the candidate's work, because this work has passed the judgment of at least a dozen different reviewers who have not seen fit to disagree with his claims, which are not necessarily mainstream within the field of neuroethics. The individual components of the framework that Zawadzki has composed are therefore already positively judged and make their author a recognized voice within the international debate on memory modification and brain interventions regarding their effects on identity and autonomy.

However, one may wonder whether there is homogeneity among the nine papers that make up the doctoral thesis and that have been published individually without a direct connection of one with the other. Each of them may be a significant and important contribution but as a whole they may not be the development of a harmonious and integrated research project on the issues of neuroethics. To carry out this assessment one must turn to the as yet unpublished part of Zawadzki's work, the section entitled "Methodology."

In it, although not particularly long, there is an interesting discussion of the characteristics that should enable neuroethics to stand as an autonomous discipline with its own status and autonomy. On this part I would like to dwell for a ideal, dialectical exchange with the author. Zawadzki states that what is peculiar to neuroethics is the neuroscience of ethics, one of the two parts into which Adina Roskies divided the new discipline in 2002. As is well known, the other part is the ethics of neuroscience. This partition has become established as a classic reference for neuroethics. I agree that I can be a good starting point, but I would tend not to absolutize it.

In addition, I believe that the ethics of neuroscience part can be an autonomous subfield of neuroethics not so much as a duplication of bioethics but as an applied ethics of a special kind, requiring a special scientific competence (think of the case of

brain organoids) combined with an ethical-philosophical background. In this sense, in contrast to Zawadzki's proposal, it is possible to think about the specificity of neuroethics not only from the side of method, but also from the side of professional figures who can be said to be neuroethicists. In this sense, it is possible to distinguish between full-time neuroethicists (scholars who devote their entire activity to teaching and scholarly production in the field of neuroethics) and intersectional neuroethicists (scholars who, while not devoting themselves completely to neuroethics, for reasons pertaining to their sphere of interests, on some occasions acquire dual expertise, for example, as a jurist who studies disorders of consciousness in order to adjudicate legal cases concerning patients with neurological deficits).

Beyond this possible expansion of criteria for establishing disciplinary boundaries, the candidate's proposal is relevant and well-motivated. His methodological line fits very well with the chosen topics, namely the concepts of personality, identity, authenticity, autonomy and agency. In this sense, specific facts can influence and inform normative thinking, based on neuroscientific and psychological findings.

The key point raised by Zawadzki concerns Hume's distinction between facts and values, the impossibility of deriving prescriptions from a factual description. If there is no relationship between facts and values, one might argue, it is useless to know more about facts in order to draw axiological conclusions from them. One must therefore admit a third way that might allow the neuroscience (or cognitive science, as Zawadzki prefers) of ethics to have a space of its own. The candidate takes this third way following Northoff's proposal that norms and facts are somehow different (whatever metaethical framework is endorsed), but there is also a principled possibility of linking normative and descriptive dimensions and thus norms and facts.

This approach works well by hooking into the above-mentioned concepts that are not in themselves immediately normative, but have an important ethical fallout and in fact are loaded in human interactions with normative value, think for example of identity. What should then be made clearer - I think the candidate will be able to do this in his or her future work - is that there are two possible normative levels - distinguishable analytically - on which to work. The first is given precisely by "bridge" concepts that are in a sense objective: they are in fact described in a scientific way by the disciplines that deal with them, think of personality, identity, agency, free will. They are studied empirically, and various theories are commercially available.

Such concepts have, as mentioned, a bridging role by also being de facto value-laden and often assuming a normative role in human interactions. Without free will there is no accountability or blame for misconduct. Thus, if neuroscience tells us that free will does not exist, the normative fallout will be very significant, because the possibility of assigning blame to an individual who violates societal norms, and thus

also of punishing him or her, is lost. Regarding personality, in the same way, if the definition of personality and the process through which we think personality is formed changes (is it an innate given or is it shaped by the environment?) we will have different rights and different obligations in diverse situations.

The approach is thus revisionist in the sense that it introduces changes in the meaning in which those concepts are used and have value and moral implications. This has further consequences in the area of ethics. The concepts considered here are defined by the candidate as "metaphysical," and in some respects their philosophical tradition certainly places them in the metaphysical domain. However, this definition may also be misleading. In fact, their possibility of being subject to a science-based revisionist approach would make them better placed in mixed theoretical-experimental realm.

In my opinion, all the concepts considered, from identity to authenticity, from autonomy to personality, have always been loaded with normative significance and have been changing their content to some extent over time. I don't think there are really any concepts that are totally unchanging and unchanged in the long run. Therefore, the idea that such concepts must be grounded in a metaphysics that is plausible in light of empirical data seems to me to address only part of the problem. In fact, the normative component always remains entrusted to intuition or argumentation with axiological premises that cannot be traced to factual aspects. In other words, the idea of authenticity, for instance, may be considered a concept that is the result of an adaptive function (it allows others to predict our behavior within a certain range), but it is also something that is judged to be of value, appreciable for its own sake, a virtue of the person in its own as a form of fidelity to their inner self.

Different is the case with a higher or basic level (depending on perspective) of normative commitment, which is not directly touched by the change in scientific knowledge. If we postulate that it is the consequences of an action that determine the value of it or that human beings should always be treated as ends and never as means or that people should educate themselves progressively to benevolence and tolerance, this comes from moral intuitions or reasoning that can only partially refer to states of affairs. This does not detract from the fact that moral values and evaluations also change over time (and sometimes as a function of new factual knowledge, think of the role exerted by biology in the case of racism).

In this sense, Zawadzki's proposal can be enriched and extended, but it has the merit, even in this formulation, of showing with specific and well-analyzed situations how the idea of cognitive science of ethics has precise and relevant effects. Of the two examples brought by the candidate, the one related to memory modifications seems to me the most convincing.

He, in fact, proceeds according to a four-stage scheme. First, the concept of authenticity is operationalized with the identification of the most influential models of authenticity, including that proposed by Pugh and colleagues (2007). Second, relevant empirical evidence is identified. Specifically, the candidate works on the integrative framework of personality proposed by McAdams and Pals (2006), which distinguishes three levels of personality. Third, he addresses empirical grounding, by selecting Pugh and collegues' model as the best empirically grounded and referring its content to McAdams and Pals' integrative framework of personality.

At this point Zawadzki can conclude with *revisionism*, that is, in assuming a model of authenticity because there seem to be good empirical reasons for accepting that model. And since that model is inherently revisionist, its application may have normative consequences. In fact, in Article 6, the candidate analyzes the risks brought by the use of optogenetics in memory modifications at the level of authenticity and personality. In the article, a careful distinction is made between memory systems whose functioning can be disturbed by neuromodulation. Then, a mapping is made of the relationship between the different memory systems and the authenticity and personality models that the candidate has selected as most closely fitting the empirical data.

At that point, a neuroethical analysis can be conducted of the potential consequences of using optogenetic techniques for memory modification at the level of personality and authenticity by highlighting the radically different effects that can occur given the type and characteristics of the specific techniques used (reversible or not, for example) and the memory systems that are affected.

Regarding the anti-retributivist approach in the penal system, I think the disagreement among scholars may be greater. Indeed, the premise of this argument for a consequentialist approach is the nonexistence of free will. It is inferred not so much on the basis of albeit relevant experiments that have recently been performed in the laboratory, but on the basis of rather controversial logical-philosophical evidence. The so-called quarantine model espoused by the candidate is attracting increasing attention, and this speaks in its favor. But I personally do not think it is yet sufficiently sound, either in its premises or its provisions, to escape major criticism.

In any case, in dealing with both examples Zawadzki moves confidently within the methodological scheme he has outlined and paves the way for new and different applications of the cognitive sciences of ethics. This is the case with Deep Brain Stimulation (DBS) whose consequences, in the candidate's opinion, do not need to be mediated through additional neuropsychological entities such as different memory systems. This is because there was direct data available, such as qualitative research on the effects of DBS on the patients. The interpretation is that neuromodulation to counteract symptoms of neurodegenerative diseases directly triggers immoral or

illegal behavior as side effects. In this case, I think instead that the interpretations may be less linear than the candidate states. How does stimulation of certain brain areas translate into specific behavior in some particular contexts? Indeed, we are not talking about continuously repeated compulsive behaviors like a tic, but complex behaviors that require some adaptation to the environment.

In this sense, one can hypothesize a slating of behavioral tendencies already present and contained by higher control mechanisms. Or one can hypothesize a complex form of stimulation whose causal pathway is not easily seen. Indeed, there do not seem to be areas deputed to every single possible immoral or illegal action that might be inadvertently stimulated by DBS. How then to deal ethically with the consequences of behavior elicitated by DBS? Can we easily solve the problem by stating that the individual was prevented from taking an alternative course of action because of deep brain stimulation? Or should we instead first identify all the intermediate brain-level hubs that lead from stimulation to the performance of a specific behavior?

These are questions that are likely still in need of extensive scientific and conceptual investigation. Exactly that investigation that has been successfully initiated by candidate Zawadzki in his doctoral dissertation. It represents an excellent beginning of a potentially very fruitful path for the development of neuroethics as an autonomous discipline, especially in its "cognitive science of ethics" part. But the 9 articles that make up the thesis also opened the way for very interesting insights in the field of memory modification, neuromodulation interventions and in the area of justification of criminal prosecution on the basis of the existence or non-existence of free will.

All of these are areas of biomedical research related to the functioning of our brains have very important repercussions on social life and the legal aspects of relationships between people. Therefore, the subject of candidate Zawadzki's investigation also makes a very important contribution to the dialogue between science, philosophy and society, which is increasingly relevant and necessary today.

I therefore rate as totally excellent the doctoral thesis done by the candidate Przemysław Zawadzki, who deserves in my opinion the highest rating provided in the Polish academic system.

Andrea Lavazza, ***University of Pavia; Centro Universitario Internazionale, Arezzo, Italy***

Milan, Italy, November 1, 2023