

Commentary on “Neuroethics at 15: The Current and Future Environment for Neuroethics”:

## **Social impact under severe uncertainty:**

### **The role of neuroethicists at the intersection of neuroscience, AI, ethics, and policy-making**

The paper “Neuroethics at 15: The Current and Future Environment for Neuroethics” by the Emerging Issues Task Force, International Neuroethics Society, addresses central challenges for neuroscience in the years to come. The authors provide examples of pressing ethical, legal, and political issues that arise from neuroscience and neurotechnology, including artificial intelligence (AI). The paper nicely illustrates how neuroscience and neurotechnology involve complex issues pertaining to epistemic uncertainty and conflicting values (for instance, between economic growth and commercial values and risks to users of the technology and to the environment). It also expresses an ambition that neuroethics should have a positive societal impact. The paper does not, however, reflect much on the proper relationship between neuroethics as academic research and the application of research and technology in real-world settings, i.e. how the translation between theory and practice should be conducted in this field of ethics (Bærøe 2014). How can, and should, neuroethicists have an impact, be policy-relevant, and inform the public? Moreover, the hope of further professionalization of neuroethics (in the Conclusion) raises the question of what constitutes expertise in this field. Does this expertise merit authority primarily in clarification and analysis of the cases at hand, or does it also give neuroethicists authority to make recommendations to policymakers and the public about what they should do?

#### *Severe uncertainty accompanying manipulated human conditions*

It is clear that neuroscience and neurotechnologies that can be used both therapeutically to restore or establish human health and functioning and to enhance human abilities beyond the frames of the human bodily design raise a variety of ethical issues. One particular challenge to which we wish to draw attention cuts to the core of what the academic field of neuroethics needs be self-reflective about. The new technological possibilities not only affect ethical issues pertaining to human conditions as we know them (by, for instance, threatening privacy, autonomy, benevolence, and justice), but may also revoke the conditions for ethics itself,

deeply challenging the pillars of social interaction and the institutions for handling normative concerns that we have leaned upon so far. For instance, replacing human intelligence (HI) with AI provides individuals with a decision-making capacity that transcends their original opportunities can be created based on algorithms developed by others. This scenario challenges and dissolves the general, default conceptualization of individuals who bear responsibility for the actions they perform—a conceptualization that, in turn, constitutes our social institutions of praise, blame, and deciding who needs special protection and who does not.

Human performance that is manipulated beyond what we could, in principle, be able to do as not-manipulated humans also raises questions concerning the justification for human equality and the equal entitlement to make decisions (at least insofar as not all are manipulated). Thus, crucial assumptions for broadly valued ideals about democracies (i.e. states being governed by the people) are also challenged. Moreover, we will argue that, if citizens are increasingly forced to depend on AI rather than their own HI for organizing their everyday professional, social, and personal lives, it is not only new conceptualizations of responsibility that are required—we may no longer be able to trust HI-based, democratic deliberation to control the further development of AI. Furthermore, when physical human presences (involving HI) are detached from the tasks of teaching, treating, and caring for fellow humans, then hands-on experience with social issues is also removed. This means that experiences that might otherwise spark compassion, solidarity, or a sense of injustice—that is, the elements in the “social glue” that essentially frames practical, real-world ethics, political ideologies, motivation, and actions—become less a part of human interaction.

We can really only guess, of course, exactly how neuroscience will change the fundamental conditions for practical ethics. Nevertheless, ethicists focusing on neuroscience in the intersecting area of AI face *epistemic uncertainty* relating to how the social impact will unfold. This includes the impact on the fundamental conditions for practical ethics as well. Individual ethicists’ attitudes toward the possibility of revoking conditions for practical ethics can take various forms, including pro-active, laissez faire, precautionary, or strongly resistant, to mention only some attitudes along a spectrum. Where in this spectrum one is positioned—knowingly or not—can affect the scope and content of concerns one considers necessary to include in an adequate ethical analysis of the acceptability of introducing new technologies in the first place. It is hard to see how the field of neuroethics can promote a unifying

*professionalism* unless a particular stance toward this pressing meta-ethical condition is collectively agreed upon. Alternatively, a shared way to deal with diverse meta-ethical opinions on this fundamental matter may need to be collectively accepted.

### *Neuroethics and expertise*

The question of how ethicists should address the overall epistemic uncertainty constitutes an area of *ethical uncertainty* in its own right, due to potentially diverging value commitments among ethicists, among the public, and between the two. On the one hand, the social impact of research is often unintended and unpredictable, and can happen without any active contribution on behalf of researcher. The role of scientists as experts in policy-making, on the other hand, involves the active contribution of scientists, which is directly coupled to practical contexts with the aim of improving public and private decision-making. Like any other branch of science, neuroethicists can, then, contribute with their knowledge by taking on the role of experts in unelected bodies, such as panels, committees, boards, and in the public sphere (for an interesting discussion of the expert role in the case of bioethics, see Lewens [2019]). This raises an important question of how neuroethicists should understand their role as experts and how an academic career can contribute to this. We will claim that, to preserve the “ethics of neuroethicists” in this role, it is crucial to include transparency regarding one’s attitude toward the fundamental changes for ethics that are potentially brought about by new technologies.

There are different ways in which neuroethicists can take on the role of expert. We can distinguish between three modes of output that experts can provide (Gundersen 2018): i) *inform* policymakers, industry, and the public by providing analysis, clarifications, and empirical claims; ii) *describe the available and feasible policies* that decision-makers can choose from (cf. Pielke’s [2007] notion of honest brokering); and iii) *make policy recommendations*. Ethicists can also *facilitate ethically justified decision-making processes* concerning design, development, implementation, and evaluation of neurotechnologies and *clarify meta-ethical conditions* (Bærøe 2014). In light of the epistemic uncertainty and lack of consensus on values, within both the community of neuroethics and the public, it is far from obvious what kind of authority neuroethicists ought to have as experts. In line with legitimate priority-setting, we would claim it presupposes, at a minimum, transparency of, and accountability for, a reasonable scope and content of considerations to include in the ethical assessment (Daniels and Sabin 2002). This brings us back to the call for stating one’s

influencing attitude toward the risk of new technologies for changing the fundamental conditions for ethical practice.

Obtaining authority as neuroethicist experts may prove challenging, since people are likely to have different opinions on how much fundamental change they are happy to welcome. This means that neuroethicists who assess the acceptability of neurotechnologies might need to take on the role of activist in order to have social impact, since leaning on the status as an accountable “expert” may not be enough. Draper (2019) has recently provided a taxonomy of various versions of activism considered relevant for academics who seek to make an impact with their work in bioethics. Activism can span from pure philosophy, theoretical applied ethics, and writing for the general public in plain English at one pole to vested interest activism and extreme vested interest activism at the other (p. 4). Ethicists working at the intersection between neuroscience, AI, and policy recommendations, in a field with multiple interests at play beyond ethics (e.g. economic, scientific, political) and without a professional consensus on the scope of acceptable impact on the conditions for ethics, would have to gravitate around the active intervention part of the continuum in Draper’s taxonomy in order for their academic work to have an impact.

### *Concluding remarks*

How can academic work in neuroethics have a social impact? To put it slightly differently, how can theoretical academic neuroethics translate into real-world, ethically justified practice? In our view, it seems reasonable that members of the neuroscientific community engage in a discussion about the nature of their expertise, how they can have a positive impact on practical decisions, and to what extent they should make ethical and political value judgments when taking on the role of experts. A careful and comprehensive discussion of how this *can* be carried out, according to the various different translational movements crossing the theory and practice divide, could serve as a collective, reflective exercise regarding the meta-ethical conditions for the field (Bærøe 2014). Moreover, the question of how this *should* be carried out can be considered essential food-for-thought for neuroethicists in the years to come.

### References

- Bærøe, Kristine. 2014. Translational ethics: An analytical framework of translational movements between theory and practice and a sketch of a comprehensive approach. *BMC Medical Ethics* 15(1): 71.
- Daniels, N., and J. E. Sabin. 2002. *Setting limits fairly: Can we learn to share medical resources?* Oxford: Oxford University Press.
- Draper, Heather. 2019. Activism, bioethics and academic research. *Bioethics*. Published ahead of print, April 25, 2019.
- Gundersen, Torbjørn. 2018. Scientists as experts: A distinct role? *Studies in History and Philosophy of Science Part A* 69: 52–9.
- Lewens, Tim. 2019. The division of advisory labour: The case of ‘mitochondrial donation’. *European Journal for Philosophy of Science* 9(1): 10.
- Pielke, Roger A., Jr. 2007. *The honest broker: Making sense of science in policy and politics*. Cambridge: Cambridge University Press.