4-1-2015

An Ethics Toolbox for Neurotechnology

Martha J. Farah
University of Pennsylvania, mfarah@psych.upenn.edu

An Ethics Toolbox for Neurotechnology

Abstract
Advances in neurotechnology will raise new ethical dilemmas, to which scientists and the rest of society must respond. Here I present a "toolbox" of concepts to help us analyze these issues and communicate with each other about them across differences of ethical intuition.

Keywords
neuroethics, bioethics, personhood, enhancement, brain-computer interface, brain stimulation, organoids, humanized animal, neural stem cell, transhumanism

Disciplines
Bioethics and Medical Ethics | Neuroscience and Neurobiology | Neurosciences

An ethics toolbox for neurotechnology

Neuron NeuroView

Martha J. Farah
Center for Neuroscience & Society
University of Pennsylvania

Author contact information:
mfarah@neuroethics.upenn.edu
215.898.1982 (fax)

Abstract

Advances in neurotechnology will raise new ethical dilemmas, to which scientists and the rest of society must respond. Here I present a "toolbox" of concepts to help us analyze these issues and communicate with each other about them across differences of ethical intuition.

Keywords: Neuroethics, bioethics, personhood, enhancement, brain-computer interface, brain stimulation, organoids, humanized animal, neural stem cell, transhumanism

It is a truism that science is a double-edged sword. 20th Century atomic physics revolutionized our understanding of material world and gave us new forms of energy, but also created the deadliest weapons of all time, which continue to threaten civilization. The 21st century's most transformative science may well be neuroscience. We are living in a time of rapid progress, as neuroscientists gain new insights into the basic science of brain function and leverage them with a range of technologies from nanomaterials to machine learning. The articles in this issue of *Neuron* show the promise held by many of these methods for advancing basic science and treating neurological and psychiatric illness.

In the midst of this rapid progress, how can we encourage the development of ethical technologies and applications? Of course we will not have complete control over the field's development, and we will not even all agree on what constitutes an ethical use. Here I suggest that a constructive first step is to stock our ethics "toolbox." These tools will help us recognize ethical issues, analyze them, and communicate with each other about them.

TWO KINDS OF TOOL: CONSEQUENTIALIST AND DEONTOLOGICAL

The ethics toolbox presented here has two main compartments, for consequentialist and deontological tools. Consequentialism is the ethical framework most closely associated with philosophers Jeremy Bentham and John Stuart Mill, according to which an act can be judged right or wrong depending on the expected value of its outcomes. Talk of "risk-benefit ratios," common in IRB applications, is a use of consequentialist ethical reasoning.

This weighing of expected value is such a natural and obvious way of approaching ethical decisions in science and technology that it may seem pedantic to give it an "ism" name and cite its 19th century roots. Indeed, I recall a bioethics meeting at which a Nobel laureate scientist impatiently asked "what's all this talk about? Just assess the benefits to patients against the risks and costs, and you'll know the right thing to do." But as this brilliant scientist came to appreciate, consequentialism alone does not fully accord with our ethical intuitions. For example, we would not be okay with sacrificing a healthy person to provide five lifesaving organ transplants, even though it seems right based on a simple calculation of aggregate benefits and costs.

The other widely used approach, which captures our sense of the wrongness of using a human being as an involuntary organ donor, is deontology, often associated with the 18th philosopher Immanuel Kant. The name 'deontology' comes from the Greek word for 'duty,' and the approach determines what is ethical in relation to a set of moral principles that specify our duties and rights as persons. Our IRBs apply such principles as well as risk-benefit calculations. For example, even if risks are negligible and benefits are substantial it would be a violation of a subject's right to autonomy to be enrolled as a research subject without informed consent.

Philosophers have attempted to reconcile the two approaches, for example by considering the beneficial consequences of recognizing rights. This has never worked satisfactorily and so we are left with fundamentally different ethical systems. For many dilemmas the same decision is recommended regardless of which ethical system we use, but conflict can arise. Indeed, there are even cases in which different deontological principles conflict with one another or different ways of weighing consequences lead to different conclusions.

What this means for the toolbox offered here is that it cannot be applied algorithmically to reach a determinate answer. What it can do is capture and highlight morally relevant considerations in a given situation, to make more explicit the grounds for various ethical positions and to facilitate discussion when disagreement occurs.

The deontology compartment: Principles for ethical decision-making

Personhood. We all share an intuition that certain entities, including ourselves, are *persons* and hence have rights and duties, whereas others, including our furniture, are not and do not. These rights and duties are spelled out in the principles of deontological ethics. Many issues in bioethics have been analyzed in terms of personhood rights. For example, if a fetus is a person then it has a right to life and abortion is wrong.

What is a person? For Kant personhood was related to the cognitive wherewithal (or cognitive potential, for the immature) to think and act morally. Others have used broader criteria, such as rationality and self-consciousness, but bioethics still lacks explicit criteria that capture our intuitions about who or what is a person without being circular (Farah & Heberlein, 2007).

Dignity. This concept was introduced into ethics by Kant as part of his explanation of how persons differ from objects. In Kant's terms, objects have prices, such that one thing can be fairly replaced by another when the prices are equal. This is not true of persons; you would not entertain a trade for friends or family members regardless of the outstanding objective traits the proposed replacement has. Persons have what Kant called a "worth beyond value," which he termed *dignity*. Recently this term has been used in a related sense by socially conservative bioethicists (Pellegrino et al., 2009) to encompass a kind of deep appreciation of humanity in all its imperfection, and has thus figured in arguments against neurotechnological enhancement of humans.

Commodification. This concept refers to the extension of market value to parts of persons and their capabilities, including organs, reproductive capabilities and cognitive capabilities.

Rights. These are moral entitlements, "must-haves" rather than "nice-to-haves;" in the words of the US Declaration of Independence, "inalienable" from persons. An example is the *right to privacy.*

Beauchamp and Childress Principles of Bioethics. Bioethicists Thomas Beauchamp and James Childress crafted a set of specific principles to guide biomedical research and practice (2012). They are: *Respect for Autonomy*, which emphasizes the right to control our own lives, *Beneficence*, which refers to the duty to help others, *Nonmalfesience*, the duty to "do no harm," and *Justice*, which concerns broader duties to society, for example promoting fairness and following the law.

Other commonly invoked principles. The toolbox has many special-purpose tools, in the form of ethical principles that capture ethical intuitions in very specific contexts. Among these are the *wisdom of repugnance*, *natural is good*, and the *therapy-enhancement distinction,* which will be explained as they become relevant later.

The consequentialism compartment: Parsing consequences for ethical decision-making

The basic tools of consequentialism are fewer and simpler than deontology (see Holland, 2003, and the Stanford Encyclopedia of Philosophy to learn more). Here I will present a few concepts that are helpful in applying consequentialism to neurotechnology.

Kinds of consequences. In Bentham's original *hedonistic consequentialism*, ethical actions are those that maximize everyone's pleasure. Because this seems to make the nucleus accumbens the arbiter of too much, a common variant is *preference or desire consequentialism*, where we act to maximize fulfillment of our more considered preferences. Of course our preferences themselves might be mistaken, so other approaches have been considered, including *perfectionist consequentialism*, which tells us to maximize the perfection or full flourishing of human potential.

Interests. These can be viewed as the consequentialist counterpart to rights, missing the obligatory nature of rights. They can be weighed relative to one another.

Externalities. Economists coined this term, referring to the effects of actions by one party on others who are not directly involved. This broadens the range of possible consequences that must be considered.

Sentience. To have interests, and thus figure in the consequentialist calculus, an entity must be *sentient*, that is, capable of experiencing perceptual and affective states. Humans are highly sentient, but at least some and perhaps all animals would also appear to be sentient.

APPLYING THE TOOLS TO NEUROTECHNOLOGY

Research Ethics

Human subjects. A mix of consequentialist and deontological considerations guide our treatment of human subjects, including risk-benefit ratio and informed consent, the latter respecting subject *autonomy*. In research with neurological or neuropsychiatric patients, subjects may lack the competence needed for informed consent, and regulations then focus on protecting the person from harm, with *nonmalfesience* a particularly important principle.

Animal subjects. The ethics of animal research is generally understood in *consequentialist* terms. Animals are viewed as *sentient* and we therefore strive to protect their *interests* as much as possible while accomplishing worthy research. The 3Rs of humane animal research (Russell & Burch, 1959) are a consequentialist amelioration of the ethical downside of animal research, based on a quantitative approach to degrees of goodness and badness. Animal research in neuroscience may be more ethically freighted than other fields, at least for modeling disorders of emotion and pain. Also relevant to the consequentialist calculation on the benefits side, the validity and usefulness of some animal models have been questioned, (Nestler & Hyman, 2010). The idea of personhood and rights for some animals is an idea with some adherents (eg, Regan, 1983).

Fetal and embryonic stem cells. Those in favor of human fetal and embryonic stem cell research typically offer consequentialist arguments about the promise of these methods for curing disease. A deontological ethical analysis will depend mainly on whether fetuses and embryos are considered persons. If they are persons, then they have a *right* to life. Even if their fate would otherwise be the medical waste container, one would be *commodifying*

them, or the reproductive functions of the parents, by using them. Although this is not my personal view, I think it is worth seeing that these objections arise from an approach to ethics that most of us have some sympathy for, even if we ultimately come down on the side of pursuing helpful new therapies.

Humanized animals. Human genes and cells can be introduced into animal brains to create human disease models. Psychological changes can likely be induced by humanization; after all, behaviors can be transferred across nonhuman species (Balaban, 2005). Given how little we know about the likely psychology of nonhuman animals that have been humanized, it may be challenging to assess their levels of comfort, suffering or frustration. This makes consequentialist analyses difficult to carry out. The primary deontological issue is which side of the person/nonperson line humanized animals are on. Sufficiently humanized primates might acquire mental capacities associated with highly developed sentience or even personhood (Greene et al., 2005). *Repugnance* is one motivator of ethical discussion of humanized animals, spurring us to question or, some might wish, limit the use of these methods.

Brains in dishes. Human neurons have been cultured in substrates that allow them to grow and connect in three dimensions, enabling the clumps of tissue to self-organize in surprisingly complex ways (Lancaster et al., 2013). How these organoids develop and what functions they might eventually display have yet to be fully grasped. They are mentioned here as a reminder of the rapidly changing nature of neuroscience research, and the possibility that radically new ethical challenges could arise in future. We can be confident that current cerebral organoids are tissue rather than brains, let alone animals, humans or persons. One would not attribute *interests*, let alone *rights,* to them. But if they become larger and more organized, is there a point at which that might change? *Repugnance*, whether "wise" or not, may give us pause about the development of such entities and invites further analysis.

Clinical applications of neurotechnology

Tissue grafting, brain computer interfaces with external or implanted components, deep brain stimulation and noninvasive brain stimulation are all in clinical use or clinical trials. Some of the ethical issues they raise are familiar from other biomedical technologies, but others relate more specifically to the novel ways in which the technologies affect the brain and call for new ethical analyses.

One such issue concerns *privacy*. The bioethics of privacy has traditionally concerned medical records. A novel issue that arises with brain-computer interfaces is the possibility of unauthorized access to patients' mental processes, inferred from their neural processes. This concern does not assume a quantum jump in brain decoding ability. Even existing technologies can derive psychological information from continuously recorded brain activity when correlated to situational factors. For example, brain activity while watching the evening news could be analyzed for a patient's responses to content associated with different political ideologies, personality traits and all manner of other traits and states. Our communications, shopping habits and other behaviors have already become more visible to others; our minds may also become more transparent in an era of devices that continuously read out brain activity.

The *autonomy* of persons may also face new challenges from neurotechnology. DARPA's RAM and SUBNETS programs (Underwood, 2013) indicate that the prospect of control over cognitive and emotional processes is not mere science fiction. Whether by hackers or Big Brother, delivery of stimulation to the brain has the potential to manipulate thought, feeling and behavior.

The ethical analysis of clinical neurotechnology has similar outcomes whether approaches with the tools of deontology, as above, or consequentialism. Focusing on consequences, it is clear that hijacking BCI and DBS systems could have extremely harmful consequences.

Neurotechnology beyond the lab and clinic

Other applications of neurotechnology are nonmedical, aimed at enhancing quality of life or achievement in healthy individuals. As neurotechologies are adopted more widely for therapeutic use, nontherapeutic uses will likely follow, just as they have for other medical therapies from plastic surgery to psychopharmacology. Indeed, noninvasive brain stimulation is already used by some with the goal of enhancing mood, concentration and gaming skills (Miller, 2014).

It is worth noting that some of the neurotechnologies discussed here may prove more effective than the pharmacologic methods of cognitive enhancement currently in use, for example the use of prescription stimulants by college students to increase attention (Smith & Farah, 2011). Compared to the broad modulatory effects of most neuropsychiatric drugs, precisely targeted stimulation or augmentation by sensory, motor and computational devices have the potential to more radically enhance human capabilities. It will be years, if not decades, before we know the impact of these technologies on healthy individuals and on society, but beginning an ethical analysis now will help us guide their development.

Enhancement has been regarded as ethically dubious by those who believe that *natural is good.* One problem with this principle is that many natural things are plainly bad – for example diseases – and we have no objections to intervening technologically in such cases. If we try to place reasonable limits on this principle, accepting the natural as an ideal to be surrendered only in case of medical need, we invoke another principle, namely that illness and health are distinct states and thus therapy and enhancement are also distinct. Conflicting with this are the many medical conditions that occur on a continuum with good health, such as hypertension and type 2 diabetes. If there is no objective *therapy-enhancement distinction*, then we cannot use it to determine which uses of neurotechnology are therapeutically justified and which not.

There are more fundamental ways in which neurotechnological enhancement grates against a deontological understanding of ethics. When we improve our psychological function by brain intervention, it is much like improving our car's performance by making adjustments under the hood. In both cases the goal is to improve function, and to the extent that we succeed without introducing problematic side effects we may view the *consequences* as good. But in so doing, we are treating a person – our self in the case of voluntary brain enhancement – as an object. Rather than improving our abilities through the exercise of our own *agency*, by effortful practice and the like, we are short-circuiting that agency and hence, the argument goes, diminishing our *personhood.* In the words of the President's Council on Bioethics under George W. Bush (2003) "personal achievements impersonally achieved are not truly the achievements of persons... [the problem] lies not in the fact that the assisting

drugs and devices are artifacts, but in the danger of violating or deforming the nature of human agency and the dignity of the naturally human way of activity. " A related point is that a focus on improved mental function *commodifies* human abilities.

A concern with *justice* calls attention to the likelihood that the benefits of brain enhancement will be enjoyed by the already privileged, a failure of *distributive justice*. But managed appropriately, enhancement could lessen the inequalities of genes and upbringing that gave some of us sharper eyes, higher IQs and happier temperaments than others (Dunlop & Savelesque, 2015).

From a *consequentialist* viewpoint, the ethics of enhancement might seem simple. In the words of John Harris (2007) "If it were not good for you it would not be an enhancement." Julian Savelescu (2009) unpacks this a bit further and reminds us that some enhancements might be better than others: "Enhancement can be defined as something which makes our lives better... What makes a good life is subject to discussion; *hedonists* believe it is the pursuit of *pleasure*, others believe that it is found in *desire fulfillment* and others in the *perfection of well-being*... Enhancement can help achieve the good life by providing ... qualities that increase the chances of us having a good life, such as health or intelligence."

The consequentialist calculus includes negative as well as positive consequences. In addition to the possible health risks of brain enhancement there could be neural tradeoffs whereby the enhancement of one ability comes at the cost of another (eg, Iuculano & Cohen Kadosh, 2013). There may also be *externalities* of the choice to enhance. Widespread enhancement at school or work will redefine "normal." Unenhanced workers who were once among the best performing will slip in their relative standing and what the boss expects from employees will be ratcheted up by each new enhancement that comes along. This creates pressure to enhance, which seems clear negative consequence. Thus, consequentialism does not give generic support the morality of enhancement; it will do its work case by case on the basis of the consequences.

Carried to extremes, the brain enhancements of the distant future may result in beings so different from us that one would call them "transhuman." There is no guarantee that continued enhancement will lead us all the way to transhumanism, but likewise there is no reason to assume it will not. Philosopher Nick Bostrom (2005) offers a primarily *consequentialist* argument for embracing transhumanism, emphasizing "the enormous potential for genuine improvements in human well-being and human flourishing that are attainable only via technological transformation." Others see negative *consequences*, including the possibility that humans like us may not fare well in a world with transhumans; we may be viewed as inferior life forms, much as we now regard chimpanzees, and treated accordingly (Warwick, 2003). This was echoed, in consequentialist and deontological terms, by political scientist Francis Fukuyama (2004) in nominating transhumanism as "the world's most dangerous idea:" "If we start transforming ourselves into something superior, what *rights* will these enhanced creatures claim, and what rights will they possess when compared to those left behind?"

Conclusion

Brain-computer interfaces and dignity may seem like the proverbial fish and bicycle – the latter has nothing to offer the former. But consider the new ethical dilemmas that will accompany neurotechnological developments, and the need for scientists, clinicians and the

rest of society to respond. The concepts of philosophical ethics are tools for bringing these issues into sharper focus, analyzing them according to more general and familiar considerations and communicating about them across differences of ethical intuition.

Acknowledgements

The writing of this article was supported by NIH grant R25 DA033023. The author is grateful to Autumn Fiester and Leo Katz for helpful comments on an earlier draft.

References cited

Balaban, E. (2005). Brain switching. *International Journal of Developmental Biology*, 49: 117-124 (2005)

Beauchamp, T.L. & Childress, J.F. (2012). *Principles of Biomedical Ethics*, 7th edition. New York: Oxford University Press.

Bostrom, N. (2005). A history of transhumanist thought. *Journal of Evolution and Technology*, 14, 1-25.

Dunlop, M., & Savulescu, J. (2015). Distributive justice and cognitive enhancement in lower, normal intelligence. *Monash Bioethics Review*, 1-16.

Farah, M.J. & Heberlein, A.S. (2007). Personhood and neuroscience: Naturalizing or nihilating? *American Journal of Bioethics – Neuroscience,* 7, 37-48.

Fukuyama, F. (2004). Transhumanism. *Foreign Policy*, No. 144, pp. *42-43*.

Greene, M., Schill, K., Takahashi, S., Bateman-House, A., Beauchamp, T., Bok, H., ... & Faden, R. (2005). Moral issues of human-nonhuman primate neural grafting. Science*,* 309, 385 – 386.

President's Council on Bioethics (2003). Beyond Therapy: Biotechnology and the Pursuit of Happiness; available at https://bioethicsarchive.georgetown.edu/pcbe/reports/beyondtherapy/index.html retrieved March 20, 2015.

Harris, J. (2007). *Enhancing Evolution: The Ethical Case for Making Better People*. Princeton University Press.

Iuculano, T. & Cohen Kadosh, R. (2013). The Mental Cost of Cognitive Enhancement. *The Journal of Neuroscience,* 33(10): 4482-4486.

Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., ... & Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature*, 501(7467), 373-379.

Miller, G. (2014, May 5). Inside the Strange New World of DIY Brain Stimulation. *Wired,* http://www.wired.com/2014/05/diy-brain-stimulation/; retrieved March 20, 2015.

Nestler, E. J., & Hyman, S. E. (2010). Animal models of neuropsychiatric disorders. *Nature neuroscience*, *13*(10), 1161-1169.

Pellegrino, E.D., Schulman, A. & Merrill, T.W. (2009). *Human Dignity and Bioethics*. Notre Dame, IN: University of Notre Dame Press.

Regan, T. (1983). *The Case for Animal Rights*. University of California Press

Russell, W.M.S. and Burch, R.L., (1959). *The Principles of Humane Experimental Technique*, Methuen, London.

Savelescu, J. (2009). Enhancement and fairness. In Healey, P. & Rayner, S. (Eds.) *Unnatural selection: the challenges of engineering tomorrow's people*. Earthscan.

Smith, M.E. & Farah, M.J. (2011). Are prescription stimulants "smart pills"? The epidemiology and cognitive neuroscience of prescription stimulant use by normal healthy individuals. *Psychological Bulletin*, 137, 717-741.

Underwood, E. (2013). DARPA Aims to Rebuild Brains. *Science*, *342*(6162), 1029-1030.

Warwick, K. (2003). Cyborg morals, cyborg values, cyborg ethics - *Ethics and Information Technology*. 5, 131-137.

Recommended for further reading:

Farah, M.J. (2010). *Neuroethics: An Introduction with Readings*. Cambridge: MIT Press.

Holland, S. (2003) *Bioethics, A Philosophical Introduction*. Oxford: Blackwell Publishing.

Zalta, E. N. (Ed.) *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/