

A helmet containing a brain–computer interface that enables the wearer to select symbols on a screen using brain activity.

## NEUROETHICS

# Agency and the algorithm

*Technologies that integrate the brain with computers have been helping people for decades. As these systems become more complex, so do the ethical issues that surround their use.*

BY LIAM DREW

“It becomes part of you,” Patient 6 said, describing the technology that enabled her, after 45 years of severe epilepsy, to halt her disabling seizures. Electrodes had been implanted on the surface of her brain that would send a signal to a hand-held device when they detected signs of impending epileptic activity. On hearing a warning from the device, Patient 6 knew to take a dose of medication to halt the coming seizure.

“You grow gradually into it and get used to it, so it then becomes a part of every day,” she told Frederic Gilbert, an ethicist who studies brain–computer interfaces (BCIs) at the University of Tasmania in Hobart, Australia. “It became me,” she said.

Gilbert was interviewing six people who

had participated in the first clinical trial of a predictive BCI to help understand how living with a computer that monitors brain activity directly affects individuals psychologically (F. Gilbert *et al. J. Sci. Eng. Ethics* 25, 83–96; 2019). Patient 6’s experience was extreme: Gilbert describes her relationship with her BCI as a “radical symbiosis”.

Symbiosis is a term, borrowed from ecology, that means an intimate co-existence of two species for mutual advantage. As technologists work towards directly connecting the human brain to computers, it is increasingly being used to describe humans’ potential relationship with artificial intelligence.

Interface technologies are divided into those that ‘read’ the brain to record brain activity and decode its meaning, and those that ‘write’ to the brain to manipulate activity in specific

regions and affect their function.

Commercial research is opaque, but scientists at social-media platform Facebook are known to be pursuing brain-reading techniques for use in headsets that would convert users’ brain activity into text. And neurotechnology companies such as Kernel in Los Angeles, California, and Neuralink, founded by Elon Musk in San Francisco, California, predict bidirectional coupling in which computers respond to people’s brain activity and insert information into their neural circuitry.

This work is being watched keenly by researchers in neuroethics — a subfield of bioethics that has emerged in the past 15 years to ensure that technologies that directly affect the brain are developed in an ethical manner.

“We don’t want to be the watchdog of neuroscience or to police how neurotechnology

should be developed,” says neuroethicist Marcello Ienca at the Swiss Federal Institute of Technology in Zurich. Instead, those in the field want to see ethics integrated into the initial design and development stages of such technologies, to maximize their benefit and to identify and minimize their potential harm — whether to individuals or to wider society.

Neuroethicists have an increasingly well-established presence in clinical settings, where they work with scientists, engineers and doctors who are developing technological approaches to treating neuropsychiatric diseases. They are following closely the evolving use of electrodes that are implanted in the brain to manipulate neural activity — a basic form of brain-writing technology — to quell the manifestations of conditions such as Parkinson’s disease and epilepsy. They are also working in laboratories that are developing brain-reading technologies to enable people who are paralysed to control prosthetic limbs and to generate speech by thought.

Already, it is clear that melding digital technologies with human brains can have provocative effects, not least on people’s agency — their ability to act freely and according to their own choices. Although neuroethicists’ priority is to optimize medical practice, their observations also shape the debate about the development of commercial neurotechnologies.

### CHANGING MINDS

In the late 1980s, scientists in France inserted electrodes into the brains of people with advanced Parkinson’s disease. They aimed to pass electrical currents through regions that they thought were causing tremors, to suppress local neural activity. This deep-brain stimulation (DBS) could be arrestingly effective: violent, debilitating tremors often subside the moment that the electrodes are activated.

The US Food and Drug Administration approved the use of DBS in people with Parkinson’s disease in 1997. Since then, the technology has come to be used in other conditions: DBS has been approved to treat obsessive compulsive disorder and epilepsy, and is being investigated for use in mental-health conditions such as depression and anorexia.

Because it is a technology that can powerfully change activity in the organ that generates our sense of personhood, DBS elicits concerns that other treatments do not. “It raises questions about autonomy because it’s directly modulating the brain,” says Hannah Maslen, a neuroethicist at the University of Oxford, UK.

Reports have surfaced about a minority of people who undergo DBS for Parkinson’s disease becoming hypersexual, or developing other impulse-control issues. One person with chronic pain became deeply apathetic after DBS treatment. “DBS is very effective,” Gilbert says, “to the point that it can distort patients’ perceptions of themselves.” Some people who received DBS for depression or obsessive compulsive disorder reported that

their sense of agency had become confused (E. Klein *et al.* *Brain-Computer Interfaces* 3, 140–148; 2016). “You just wonder how much is you anymore,” said one. “How much of it is my thought pattern? How would I deal with this if I didn’t have the stimulation system? You kind of feel artificial.”

Neuroethicists began to note the complex nature of the therapy’s side effects. “Some effects that might be described as personality changes are more problematic than others,” says Maslen. A crucial question is whether the person who is undergoing stimulation can reflect on how they have changed. Gilbert, for instance, describes a DBS patient who started to gamble compulsively, blowing his family’s savings and seeming not to care. He could only understand how problematic his behaviour was when the stimulation was turned off.

Such cases present serious questions about how the technology might affect a person’s ability to give consent to be treated, or for treatment to continue. If the person who is undergoing DBS is happy to continue, should a concerned family member or doctor be able to overrule them? If someone other than the patient can terminate treatment against the patient’s wishes, it implies that the technology degrades people’s ability to make decisions for themselves. It suggests that if a person thinks in a certain way only when an electrical current alters their brain activity, then those thoughts do not reflect an authentic self.

Such dilemmas are thorniest under conditions in which the explicit goal of treatment is to change traits or behaviours that contribute to a person’s sense of identity, such as those associated with the mental-health condition anorexia nervosa. “If, before DBS, a patient says, ‘I’m somebody who values being thin over all other things,’ and then you stimulate them and their behaviour or outlook is modified,” Maslen says, “it’s important to know whether such changes are endorsed by the patient.”

She suggests that when the changes align with therapeutic objectives, “It is perfectly coherent that a patient could be happy with the ways in which DBS changes them.” She and other researchers are working to design better consent protocols for DBS, including extensive consultations in which all possible outcomes and side effects are explored in depth.

### READING THE BRAIN

To observe a person with tetraplegia bringing a drink to their mouth using a BCI-controlled robotic arm is spectacular. This rapidly advancing technology works by implanting an array of electrodes either on or in a person’s motor cortex — a brain region involved in planning and executing movements. The activity of the

brain is recorded while the individual engages in cognitive tasks, such as imagining that they are moving their hand, and these recordings are used to command the robotic limb.

If neuroscientists could unambiguously discern a person’s intentions from the chattering electrical activity that they record in the brain, and then see that it matched the robotic arm’s actions, ethical concerns would be minimized. But this is not the case. The neural correlates of psychological phenomena are inexact and poorly understood, which means that signals from the brain are increasingly being processed by artificial intelligence (AI) software before reaching prostheses.

Philipp Kellmeyer, a neurologist and neuroethicist at the University of Freiburg, Germany, says that applying AI and machine-learning algorithms to analysing and decoding neural activity has “turbocharged the whole field”. He highlights work, published in April, in which such software interpreted neural activity that occurred while people with epilepsy silently mouthed words, and then used this information to generate synthetic speech sounds (G. K. Anumanchipalli *et al.* *Nature* 568, 493–498; 2019). “Two or three years ago,” he says, “we’d have said either that would never be possible, or it was at least 20 years away.”

But, he says, using AI tools also introduces ethical issues of which regulators have little experience. Machine-learning software learns to analyse data by generating algorithms that cannot be predicted and that are difficult, or impossible, to comprehend. This introduces an unknown and perhaps unaccountable process between a person’s thoughts and the technology that is acting on their behalf.

Developers are realizing that prostheses work more efficiently when certain computations are left to BCI devices, and when these devices try to predict what the user will do next. The benefits of offloading computations are obvious. Seemingly simple acts such as picking up a cup of coffee are actually highly complex: people subconsciously execute many computations. Fitting prostheses with sensors and mechanisms for autonomously generating coherent movements makes it easier for users to perform tasks. But this also means that much of what robotic limbs do is not actually directed by the user.

The predictive nature of some algorithms used to help people operate prostheses leads to further concerns. Predictive text generators that are found in mobile phones highlight this issue: they can be useful, time-saving tools, but anyone who has sent an unintended message owing to an errant auto-correct or auto-fill function knows how things can go wrong.

Such algorithms learn from previous data and guide users towards decisions on the basis of what they have done in the past. But if an algorithm constantly suggests a user’s next word or action, and the user merely approves that option, the authorship of a message or movement will become ambiguous. “At some point,” Kellmeyer says, “you have these very strange

**“If you have a device that constantly steps up in your thinking or decision-making, it might compromise you as an agent.”**



situations of shared or hybrid agency.” Part of the decision comes from the user, and part comes from the algorithm of the machine. “It opens up a problem — an accountability gap.”

Maslen is confronting this problem as part of a collaborative project called BrainCom, funded by the European Union, that is developing speech synthesizers. Such technology has to accurately vocalize what users want to say to be useful. To guard against errors, users could be given the opportunity to approve each word for broadcast — although constantly and covertly relaying speech fragments to the user for review might make for a cumbersome system.

Safeguards such as this would be especially important if devices struggled to distinguish between neural activity intended for speech and that which underlies private thought. Societal norms require that the fundamental boundary between private thought and outward behaviour be protected.

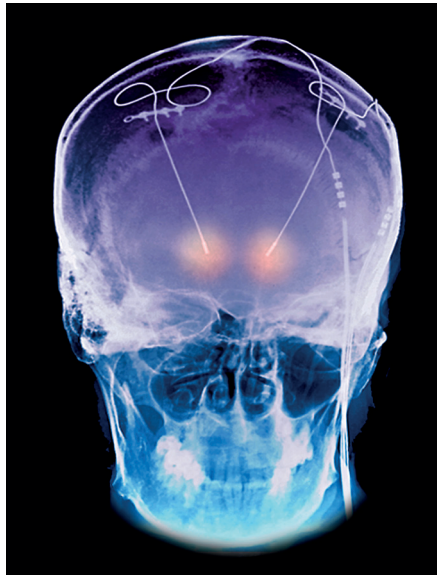
### READING, WRITING AND RESPONSIBILITY

Because the symptoms of many brain diseases come and go, brain-monitoring techniques are increasingly being used to directly control DBS electrodes so that stimulation is provided only when needed.

Recording electrodes — such as those that warned Patient 6 of impending seizures — track brain activity to determine when symptoms are happening or are about to occur. Rather than merely alerting the user to the need to take action, they trigger a stimulating electrode to nullify this activity. If a seizure is probable, DBS quietsens the causative activity; if tremor-related activity increases, DBS suppresses the underlying cause. Such a closed-loop system was approved by the Food and Drug Administration for epilepsy in 2013, and such systems for Parkinson’s disease are edging closer to the clinic.

For neuroethicists, one concern is that inserting a decision-making device into someone’s brain raises questions about whether that person remains self-governing, especially when these closed-loop systems increasingly use AI software that autonomously adapts its operations. In the case of a device for monitoring blood glucose that automatically controls insulin release to treat diabetes, such decision-making on behalf of a patient is uncontroversial. But well-intentioned interventions in the brain might not always be welcome. For instance, a person who uses a closed-loop system to manage a mood disorder could find themselves unable to have a negative emotional experience, even in a situation in which it would be considered normal, such as a funeral. “If you have a device that constantly steps up in your thinking or decision-making,” says Gilbert, “it might compromise you as an agent.”

The epilepsy-management device used by Patient 6 and the other recipients that Gilbert interviewed was designed to keep patients in control by sounding a warning about impending seizures, which enabled the patient to



Electrodes for deep brain stimulation implanted in a person who has Parkinson’s disease.

choose whether to take medication.

Despite this, for five of the six recipients, the device became a major decision-maker in their lives. One of the six typically ignored the device. Patient 6 came to accept it as an integral part of their new self, whereas three recipients, without feeling that their sense of self had been fundamentally shifted, were happy to rely on the system. However, another was plunged into depression, and reported that the BCI device “made me feel I had no control”.

“You have the ultimate decision,” Gilbert says, “but as soon as you realize the device is more effective in the specific context, you won’t even listen to your own judgement. You’ll rely on the device.”

### BEYOND THE CLINIC

The goal of neuroethicists — to maximize the benefits of emerging techniques and to minimize their harm — has long been entrenched in medical practice. The development of consumer technology, by contrast, is notoriously covert and subjected to minimal oversight.

With technology companies now investigating the feasibility of mass-market BCI devices, Ienca thinks that this is an important moment. “When a technology is in its germinal stage,” he says, “it’s very hard to predict the outcomes of that technology. But when the tech is mature — in terms of market size or deregulation — it can be too societally entrenched to improve it.” In his opinion, there is now sufficient knowledge to act in an informed manner, before neurotechnology is widely used.

One issue that Ienca is addressing is privacy. “Brain information is probably the most intimate and private of all information,” he says. Digitally stored neural data could be stolen by hackers or used inappropriately by companies to whom users grant access. Ienca says that neuroethicists’ concerns have forced developers to attend to the security of their devices, to

more diligently protect consumer data, and to cease demanding access to social-media profiles and other sources of personal information as a condition of a device’s use. Nevertheless, as consumer neurotechnology gains steam, ensuring that privacy standards are acceptable remains a challenge.

Privacy and agency feature prominently in recommendations that are being produced by various working groups, including large-scale neuroscience projects and panels convened by independent bodies. But Kellmeyer thinks that there is still considerable work to be done. “The matrix of traditional ethics, which focuses on autonomy, justice and related concepts, will not be enough,” he says. “We also need an ethics and a philosophy of human–technology interactions.” Many neuroethicists think that the ability to directly access the brain will make it necessary to update basic human rights.

Maslen is already helping to shape BCI-device regulation. She is in discussion with the European Commission about regulations it will implement in 2020 that cover non-invasive brain-modulating devices that are sold straight to consumers. Maslen became interested in the safety of these devices, which were covered by only cursory safety regulations. Although such devices are simple, they pass electrical currents through people’s scalps to modulate brain activity. Maslen found reports of them causing burns, headaches and visual disturbances. She also says clinical studies have shown that, although non-invasive electrical stimulation of the brain can enhance certain cognitive abilities, this can come at the cost of deficits in other aspects of cognition.

Maslen and her colleagues wrote a policy paper targeted at European regulators who were reviewing the regulation of various quasi-medical products such as laser hair-removal devices. The regulators agreed with the paper’s recommendations: that the new regulations should tighten safety standards, but also that (unlike for medical devices) consumers should remain free to decide whether the devices bring the gains that their manufacturers claim.

Gilbert’s continuing work on the psychological effects of BCI devices highlights the stakes that are involved in companies developing technologies that can profoundly shape a person’s life. He is now preparing a follow-up report on Patient 6. The company that implanted the device in her brain to help free her from seizures went bankrupt. The device had to be removed.

“She refused and resisted as long as she could,” says Gilbert, but ultimately it had to go. It’s a fate that has befallen participants of similar trials, including people whose depression had been relieved by DBS. Patient 6 cried as she told Gilbert about losing the device. She grieved its loss. “I lost myself,” she said.

“It was more than a device,” Gilbert says. “The company owned the existence of this new person.” ■

Liam Drew is a writer based in London.