CHAPTER 14

Ethical considerations for fMRI neurofeedback

Mouslim Cherkaoui[a], Jesse Rissman[a], Hakwan Lau[a,b], and Michelle Hampson[c]
[a]*Department of Psychology, University of California, Los Angeles (UCLA), Los Angeles, CA, United States,* [b]*State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong,* [c]*Department of Radiology and Biomedical Imaging, Yale University School of Medicine, New Haven, CT, United States*

1 Introduction

Ask students enrolled in higher education whether it would be convenient to learn a new skill unconsciously, without effort or concentration. The answer will likely be "yes." And how about erasing the memory of a difficult breakup? To some people, such possibilities might constitute a violation of their ethical principles. To others, such possibilities might be appealing.[a]

Market trends have witnessed an increased interest from the general public in technologies that claim to read brain signals or stimulate the brain. There is now a significant "do-it-yourself" movement in which consumers self-administer brain stimulation (Zuk et al., 2018). According to one estimate, the neurointervention industry is expected to reach about $13 billion in 2022.[b]

Neurointervention technologies remain controversial, as many companies make unsubstantiated claims about their products. This has led to calls for regulation and oversight (Coates McCall et al., 2019; Wexler and Thibault, 2019). Additionally, the advent of novel neurotechnologies has led to calls for reconceptualization and updating of ethics guidelines (Yuste et al., 2017) and human rights (Ienca and Andorno, 2017) to consider these emerging technologies.

Although science fiction often portrays neurotechnology as a powerful tool for enhancement of mental function (e.g., *Limitless*) and memory editing (e.g., *Eternal Sunshine of the Spotless Mind*), current tools have relatively limited capabilities (Westbrook et al., 2020). Even so, newly emerging neurotechnologies can be used to attain meaningful behavioral and neural changes. Functional magnetic resonance imaging (fMRI)-based neurofeedback is one such powerful method.[c]

a https://www.npr.org/sections/health-shots/2018/09/18/646251015/this-rapper-tried-to-use-neuroscience-to-get-over-her-ex.

b https://www.prnewswire.com/news-releases/global-neurotechnology-market-2018-2022-largest-segment-is-currently-neuromodulation-followed-by-neuroprosthetics-and-neurosensing-300775578.html.

c fMRI neurofeedback will hereby be referred to as "neurofeedback" unless stated otherwise.

fMRI Neurofeedback. https://doi.org/10.1016/B978-0-12-822421-2.00007-7

Some neurofeedback methods are particularly controversial because they are able to attain changes in brain function without the participant's knowledge of the purpose of the training or awareness that they are being trained at all (Ramot et al., 2016). *Implicit neurofeedback*, in particular, can be used to change a participant's preferences (Shibata et al., 2016) or potentially to incept memories (Amano et al., 2016; see deBettencourt and Norman (2016) for discussion on using implicit neurofeedback to implant memories for associations that were never consciously experienced). Press articles covering neurofeedback studies have noted that neurofeedback "sounds [..] like brainwashing,"[d] and that "if neurofeedback is to reach its potential, it demands we think carefully about the scientific and philosophical basis of human agency."[e] As technology advances, it is expected that this method will become more powerful. Thus neurofeedback's yet unleashed potential requires serious ethical consideration for researchers and clinicians. Indeed, the United States Presidential Commission for the Study of Bioethical Issues concluded in its 2015 report: "We should be open to the possibilities that neuroscience can bring, while ensuring the progress and responsible application of neuroscience to the legal system and policymaking." The goal of this chapter is to discuss the ethical issues surrounding the use, and potential abuse, of fMRI neurofeedback.

Some of the issues are not entirely new or unique to neurofeedback as they overlap with other types of neurointervention technologies (hereby collectively referred to as *neurotechnologies)*. Indeed, the debate as it pertains to other forms of neurotechnologies is very active. For an extended discussion of these issues, we refer readers to Linden (2014). It is not possible to do full justice to all the details and subtleties here, but our aim will be to highlight some of the main lines for argumentation as far as they are also relevant to fMRI neurofeedback. Identifying parallel problems will allow researchers and practitioners to get insight into how to best handle problems using existing theoretical resources before the problems arise, and will also help to identify issues that are unique to fMRI neurofeedback.

2 Ethical issues in common with other neurotechnologies

To understand what makes neurofeedback different from other neurotechnologies we first have to discuss what neurofeedback has in common with other neurotechnologies. Our focus here will be on ethical issues that also arise for so-called smart drugs (pharmaceuticals that putatively confer cognition-enhancing properties) and neurostimulation devices.

d https://spectrum.ieee.org/the-human-os/biomedical/imaging/dont-like-his-face-brain-train-the-feeling-to-be-more-positive.

e https://aeon.co/essays/neurofeedback-can-zap-your-fears-without-you-even-knowing.

2.1 Cognitive liberty

The notion that individuals' preferences and memories should not be altered without their consent is a central ethical concept. The right to control one's own consciousness and thought process has been argued to be fundamental to just about every other freedom (Lavazza, 2018; Sententia, 2006). Sententia (2006) defined cognitive liberty as "every person's fundamental right to think independently, to use the full spectrum of his or her mind, and to have autonomy over his or her own brain chemistry" (p. 223). Bublitz (2013, p. 251) argued that cognitive liberty requires both the liberty to change one's mind and the protection of mental integrity. Cognitive liberty is an updated notion of "freedom of thought" that takes into account the ability to monitor and manipulate cognitive function *by means other than one's own* (Bublitz, 2013; Sententia, 2006). As we will see later, what is meant by "by means other than one's own" is itself subject to a person's *perception* of their mental processes as truly being "their own."

To begin to understand the importance of cognitive liberty, let us consider a study in which researchers collected end-user data of people that had undergone deep brain stimulation (DBS) to treat their depression (Klein et al., 2016). Some of these patients reported that after surgery they had begun to wonder whether the device produced emotions artificially. The patients also reported that they worried about lost agency over their actions and whether the way they interacted with others was actually due to the DBS device. As one patient recounted "your feelings do not get attributed to you, they are all due to the device" (Klein et al., 2016, p. 5). This study showed that neurotechnology can profoundly disrupt people's sense of identity.

Of course, DBS treatments are invasive and therefore often the option of last resort for depression treatment. However, while neurofeedback is much safer from the perspective of physical risk, it shares with DBS a risk to cognitive liberty. That is, it is possible that by efficiently manipulating some emotions or preferences, or incepting memories, neurofeedback could cause participants to begin to question their autonomy and agency. It might also be possible that long-lasting mental and behavioral changes could be instilled in participants through neurofeedback. This may lead to compromised selective expression, and exemplifies that cognitive liberty may be relinquished to some extent for patients and participants undergoing neurofeedback. One review of the ethics of neurotechnologies and artificial intelligence by Yuste et al. (2017) noted that existing ethics guidelines are insufficient to address concerns such as these. Indeed, they assert that the Declaration of Helsinki, a statement of ethical principles involving human subjects, and the Belmont Report, a statement by the US National Commission of the Protection for Human Subjects of Biomedical and Behavioral Research, are both deficient in this regard.

Cognitive liberty concerns whether the participants *perceive* their own mental process as their own. The related concept of mental integrity is slightly different because it is not so much concerned with how a person perceives their mental processes. Mental integrity refers to the right of an individual's neural computation to be protected. Mental integrity is considered by the European Convention of Human Rights and the European Union Charter of Fundamental rights as a right to mental health that is understood more broadly as part of physical health by the World Health Organization. Given that the aim of neurofeedback is to alter neural and thereby mental function, issues of mental integrity and cognitive liberty are raised even when the induced alterations in mental function are desirable, as discussed next.

2.2 Enhancement

There are many neurotechnologies that aim to enhance cognition. Enhancement refers to a biomedical intervention that is meant to improve or reform the body or mind above physical or mental capacities (President's Council of Bioethics, 2003). Such technologies include noninvasive neurostimulation devices such as transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS). Noninvasive brain stimulation has been used with some success to enhance complex cognitive processes such as source memory retrieval and arithmetic ability (Snowball et al., 2013; Westphal et al., 2019). Pharmacological drugs for neuroenhancement were originally developed to ameliorate neuropsychiatric disorders but subsequently reappropriated to improve focus, attention, and emotion recognition in healthy people (Brühl et al., 2019; Buyx, 2015). Examples of these "smart drugs" include Methylphenidate (sold under the trade name Ritalin), Amphetamines (sold under the trade name Adderall, for example), and Modafinil (sold under the trade name Provigil).

Similar to these other neurotechnologies, neurofeedback has successfully been used to improve attention during a sustained attention task and to boost recall of specific positive autobiographical memories (deBettencourt et al., 2015; Young et al., 2017). While the neurofeedback effects demonstrated to date are relatively focal, in contrast to the domain general effects typically seen from smart drugs or neurostimulation, this does not mean that more pervasive effects will not prove feasible as neurofeedback development proceeds.

Although some forms of neurofeedback may eventually provide effective cognitive enhancement, there is concern that the efficacy of neurofeedback technologies that are publicly available will be overstated and marketed in a misleading way. This is a concern shared by other neurotechnologies. The potency of smart drugs and neurostimulation devices has sometimes been criticized as "hyped up" or exaggerated in the media (Brühl et al., 2019; Zuk et al., 2018). Actual evidence for positive effects of smart drugs remains mixed (Battleday and Brem, 2015; Brühl et al., 2019; Ilieva et al., 2015; Partridge et al., 2011). Nevertheless, public interest has been encouraged by media reports (Zuk et al., 2018). According to one estimate the widespread use of smart drugs is increasing (Maier et al., 2018). Overall, studies have suggested that the public view neurotechnologies both with promise and concern (Bard et al., 2018; Medaglia et al., 2019). As with other neurotechnologies, it is important that researchers engage in discussion with the media and regulatory agencies about the potential uses and misuses of neurofeedback, and work to minimize the impact of misleading media coverage. Clear communication with the public regarding the capabilities and risks of neurofeedback is critical to minimizing the potential for unethical application or marketing.

2.2.1 The enhancement/treatment and authenticity debate

One philosophical question that is relevant for all neuroenhancement is: How do we distinguish between enhancement and treatment? The assumption inherent in any answer to such a question is that a clear border *exists* between enhancement for nontherapeutic purposes and treatment of a medical condition. It is assumed that this border can be used to determine when neurotechnologies are regarded as treatment for an illness and or an intervention that is mainly for personal interest, which is not deemed to be absolutely necessary. But it is difficult to draw this border in absolute terms because it would depend on the elusive concepts of health and "disease" (Boorse, 2011; Schermer, 2015).

The President's Council on Bioethics (2003) evaluated smart drugs negatively because of fears that users would exhibit changes to their personalities and thus become less authentic. The authenticity argument states that for a person to be authentic, they have to be true to oneself and feel that their feelings are their own. By altering cognitive states through smart drugs, people may become alienated from their own experiences and feelings, and have a distorted perception of the world from what it *truly* is (President's Council, 2003). On this view, the consequence of a society-wide dissemination of smart drugs would be that by taking smart drugs as a "quick fix" one can no longer claim one's own efforts and experiences as truly *one's own* (Buyx, 2015). And some have argued that memory-editing drugs may pose a particular threat to one's authenticity by deceiving oneself about the content of past experiences and influencing our "natural disposition" to respond to life events in a certain way (Erler, 2011).

However, for some people, using enhancement neurotechnologies may allow them to *attain* a sense of authenticity because they feel a detachment between their current self and their true self. Furthermore, another counterargument to those who view enhancement technologies as a violation of authenticity is that enhancement interventions exist on a continuum starting from coffee, to exercise, to brain surgery. It is not necessarily the case that just because these interventions are also used to treat illnesses that they cannot be ethical for enhancement purposes.

2.3 Safety and efficacy

Like any medical intervention, neurofeedback, smart drugs, and neurostimulation devices have faced questions of safety and efficacy. For example, because smart drugs were originally developed for and trialed on clinical populations with diagnosed diseases, evidence is lacking on the risks of use in those that have not been diagnosed. Fortunately, due in part to its value as a basic science research tool, early development of fMRI neurofeedback has involved extensive research in cognitively healthy as well as patient populations. Including careful adverse events monitoring in these basic science neurofeedback studies, as well as the clinical studies, could ensure that knowledge regarding the safety and efficacy advance together for both patient and healthy populations.

Of particular ethical concern is the theoretical possibility that one could implement neurofeedback to *worsen* symptoms or impair cognition and behavior. Given our limited understanding of brain function, this can happen even when researchers have the best intentions. For instance, Ruiz et al. (2013) trained schizophrenia patients to achieve control of bilateral anterior insula cortex to improve facial emotion recognition (a common deficit in schizophrenia). While patients increased their ability to recognize disgust faces, their ability to recognize happy faces declined. As neurofeedback and our understanding of brain function develops, the capacity to target specific emotions and memories will increase, and accidental downstream effects will hopefully be minimized. There will, however, always be a potential for abuse. One can imagine a scenario in which neurofeedback originally developed to influence face preferences (Shibata et al., 2016) could instead be used by a malicious experimenter to try to make themselves more attractive to the subject. Similarly, neurofeedback could be used to influence preferences for specific brands of expensive products. Indeed, such a scheme could be accomplished if companies "sponsor" neuroimaging studies in return for a short neurofeedback scan that influences the subject without their knowledge. While this might seem unlikely, the research community needs to take such a possibility seriously and be prepared to thwart any insidious applications of neurofeedback technology. Doing so will involve proactively engaging with the public, regulators, and other researchers to create a dialogue around what types of interventions are considered ethically acceptable in which contexts.

One important distinction between neurofeedback and pharmacological drugs is that while the cognitive and behavioral impact of pharmacological drugs is continuously diminished by metabolic processes, neurofeedback can likely have more long-lasting effects. Recent studies have suggested that only a few neurofeedback sessions can induce effects that may last up to months or years (Megumi et al., 2015; Amano et al., 2016; Ramot et al., 2016, 2017; Rance et al., 2018; Robineau et al., 2017), but further work is needed to clarify the persistence of effects (both desired and undesired) across applications (Paret et al., 2019; Rance et al., 2018; Thibault et al., 2018). This is relevant to both the clinical efficacy of a neurofeedback-based intervention and the potential risks it engenders. More data on persistence of effects are necessary from an ethical perspective to evaluate the costs and benefits of neurofeedback interventions.

2.3.1 Enhancing the developing brain

Cohen-Kadosh et al. (2012) pointed out that great caution should be taken with regard to neurostimulation of the developing brain. Enhancing some abilities through neurostimulation may come at a cost for some other modalities, which may cause interference with normal brain development (Doruk Camsari et al., 2018; Hameed et al., 2017; Maslen et al., 2014). This reasoning has led to the suggestion that brain stimulation for enhancement purposes should be delayed until "the child has reached a state of maturity" (Maslen et al., 2014, p. 2). Similar concerns have been raised against shaping not-yet-mature brain networks of children through neurofeedback (Cohen-Kadosh et al., 2016). On the other hand, early intervention has potential to yield the greatest lifelong benefits by harnessing the high plasticity of childhood, and steering development at an early stage onto a more positive trajectory. Thus developmental applications of fMRI neurofeedback are potentially valuable but also potentially dangerous, and the need for careful monitoring of long-term outcomes is particularly acute in pediatric populations.

2.4 Reversibility of effect

The reversibility of the effect is directly related to Bublitz's first dimension of cognitive liberty, namely the liberty to change one's mind. Just as it is important to demonstrate that neurofeedback can induce an effect, it is important to determine if the effect can be suppressed or eliminated entirely with reverse feedback. This is especially relevant given reports that just a few training sessions can have lasting effects (Amano et al., 2016; Megumi et al., 2015; Ramot et al., 2017). There is also practical value in being able to "cancel out" any negative behavioral side effects of neurofeedback, should these occur.

Very little is known regarding the reversibility of neurofeedback effects. Cortese et al. (2017) demonstrated that it is possible, by training brain patterns in the opposite direction, to reverse to some extent the behavioral changes induced by neurofeedback within single participants. Participants were first trained, via decoded neurofeedback, to either up-regulate or down-regulate their level of confidence in a perceptual task, and then they were trained to modulate confidence in the opposite direction. The results showed evidence for "anterograde interference" such that previous learning interfered with subsequent learning. Indeed, once participants had been trained to up-regulate their confidence, retraining in the opposite direction only canceled out 20% of this effect, indicating that 80% of the original effect size persisted despite the experimenters' efforts to override it. Of course, the efficacy of cancellation is likely dependent on the number of induction training and reverse training days. Furthermore, reversibility is likely to be highly context dependent, varying depending on the circuitry trained and the timing between training and reverse training sessions. More studies are needed that investigate reversibility of neurofeedback effects. In particular, such work will be valuable for neurofeedback interventions that are demonstrated to be effective before they are considered/adapted for widespread application.

3 Distinct ethical issues raised by fMRI neurofeedback

In comparison with other noninvasive neurointerventions (pharmaceutical and transcranial stimulation-based approaches), neurofeedback effects on the brain can be targeted much more flexibly. That is, detailed and complex spatial patterns of brain activity (or functional connectivity) corresponding to biomarkers of the mental functions of interest can be trained to be induced by participants via fMRI neurofeedback. Although invasive neurostimulation (e.g., intracranial recordings) can achieve superior spatial specificity, fMRI neurofeedback is unique in that it may have the potential to effectively mold very precise spatial patterns of brain activity/connectivity in a completely noninvasive manner, and even without awareness of training occurring (Ramot et al., 2016; Taschereau-Dumouchel et al., 2018a). To the extent that this capacity translates into a powerful tool for mental control, existing ethical concerns may be amplified, and unique ethical issues may be raised.

For the purposes of this discussion, it is helpful to distinguish two types of ethical concerns associated with fMRI neurofeedback. One is concerned with the social impact of mental influence if exerted broadly across a large population of people. Even a small influence, of little note to an individual neurofeedback participant, could have dramatic repercussions on society if leveraged across large numbers of people. The second concerns the well-being of the trained individuals, and potential violations of their individual autonomy due to substantial influence over their personal mental, emotional, or perceptual function. These two possible concerns are described in more detail below, after which we discuss implicit neurofeedback, a form of neurofeedback that highlights these ethical challenges.

3.1 The potential for broad social impacts of neurofeedback

We believe that fMRI neurofeedback itself is extremely unlikely to be used for broad social influence, given the expense and inconvenience of MRI scanning. However, the advances it provides in terms of identifying neurofeedback targets that can be trained to influence specific aspects of mental function, if translated to other neuroimaging modalities, could pose risks of this nature. This concern is thus based heavily on speculation regarding the development and translation of neurofeedback. However, it is important to note that translation of neurofeedback to other modalities is currently a research area of great interest (Keynan et al., 2016, 2019; Meir-Hasson et al., 2014, 2016). In this current age where technical developments can sweep through society more quickly than ethical frameworks can develop, some thoughtful speculation and discussion is healthy for ensuring socially responsible development.

In a talk in 2019 at Harvard University, Mark Zuckerberg reported that Facebook is working on developing a portable headset for monitoring brain function. To quote from a *Wired* article reporting on this talk "The technology that Zuckerberg described is a shower-cap-looking device that surrounds a brain and discovers connections between particular thoughts and particular blood flows or brain activity, presumably to assist the glasses or headsets manufactured by Oculus VR, which is part of Facebook."[f] It's conceivable the public would embrace this new technology when marketed for its convenience and entertainment value. However, the linking of a device that monitors brain function to a feedback display via glasses or headsets would provide the full technological setup for neurofeedback. Successful dissemination of such a device to large numbers of people (presumably, the goal of Facebook) would create a science fiction-like scenario with frightening potential for social influence and control. Of course, the ability of the device to accurately measure complex brain patterns of relevance to mental function would be a critical limiting factor over its ability to influence people's thoughts and behaviors. However, even a small influence in a critical direction applied to large numbers of people can be very dangerous. Apparently, the device already "can distinguish when a person is thinking of a giraffe or an elephant based on neural activity."[f] In short, the potential for development along these lines is concerning. Communication with the public and government regulatory agencies regarding these risks, and the development of guidelines for preventing abuse of such technologies are needed.

3.2 Potential risks to individual subjects

The origins of neurofeedback trace back to operant autonomic conditioning. During the early days of electroencephalography (EEG) neurofeedback, many enthusiasts hailed EEG neurofeedback as a technology that could allow "mind-expansion" (Kimmel, 1986). While research in this area has progressed, in reality, EEG neurofeedback applications have to date proven more limited. It is similarly difficult to foresee how potent fMRI neurofeedback will prove to be as a tool for molding mental function. Although recent coverage has noted that fMRI neurofeedback "sounds... like brainwashing"[d] it has not been demonstrated as enabling effective control over the complex thought patterns of an individual to the degree typically associated with brainwashing, that is, to the degree that an individual's strongly held views can be reversed or specific thoughts can be introduced against their will. However, it has been demonstrated that learned associations can be trained into an individual via neurofeedback (Amano et al., 2016), that preferences can be altered (Shibata et al., 2016), and perceptions can be biased (Shibata et al., 2011). Thus fMRI neurofeedback can be used to alter some aspects of mental function, and an important question is whether further development will lead to a more powerful tool for mind control.

f https://www.wired.com/story/zuckerberg-wants-facebook-to-build-mind-reading-machine/.

To place this issue in context, it is important to note that the ability to influence mental function is not unique to neurofeedback, as basic psychological methods can also mold mental function. For example, researchers have demonstrated the ability to create a "false memory" simply by using suggestive techniques (Loftus, 2005). In one study, researchers falsely suggested to their participants that the participants had become ill after eating an egg salad as a child (Geraerts et al., 2008). This led some participants to confidently believe that they indeed had experienced this childhood event at a 4-month follow-up, even though they had previously denied having experienced it. What is more, the 4-month follow-up found that believers' eating behaviors and preferences reflected an intent to avoid egg salad. Geraerts et al. (2008) comment in their discussion that researchers should consider that they may induce lasting false beliefs in their participants that "have consequences not only for attitudes, but also behavior" (p. 752). We note that this is very similar to concerns that have been raised for neurofeedback protocols (Nakazawa et al., 2016; Sadato et al., 2019; see also Horstkötter, 2016).

Given that neurofeedback appears capable of inducing long-lasting behavioral, perceptual, and cognitive changes, researchers should give ethical consideration to the ways in which they develop and implement this technology. Important questions to answer include: Should neurofeedback be allowed to "enhance" cognitive capacity in healthy humans (e.g., to boost one's ability to flexibly shift or sustain attentional focus, maintain and manipulate information in working memory, form and retrieve long-term memories, make more rational decisions, etc.)? Should researchers be allowed to create artificial memories (e.g., to help people replace fearful associations with safer ones) and change preferences (e.g., to promote healthier eating habits or diminish harmful addictions)? Should researchers be allowed to manipulate emotions, and if so, which ones and in what contexts (e.g., to reduce sadness in depressed patients or worry in patient's anxiety disorders)? What are the long-term effects of such manipulations? What could constitute an adverse effect? What impacts could wide dissemination of neurofeedback have on societal norms and values? And how could neurofeedback be abused in the wrong hands? These are all important ethical questions that have yet to be fully explored and answered for modern fMRI neurofeedback.

3.3 Implicit neurofeedback

As discussed before, some forms of neurofeedback involve keeping the participant unaware of the precise nature of the intervention (e.g., Shibata et al., 2011). Implicit neurofeedback studies are important because they prevent participants from knowing the experimenters' goals and having expectations regarding the outcome of the training. In this way, implicit neurofeedback protocols can avoid potential confounds such as demand bias (participants reporting what they think the experimenter is hoping they will report) and placebo effects (participants experiencing effects consistent with their expectations). Theoretically, implicit neurofeedback may also be more efficient for some applications, by allowing subconscious learning to proceed without the interference of conscious effort (see, e.g., Sepulveda et al., 2016). Finally, implicit neurofeedback can be used in populations that cannot follow complex instructions or as an alternative for populations that find explicit therapeutic training aversive (Watanabe et al., 2017).

Implicit protocols have been demonstrated to induce effects with high degrees of behavioral and neural specificity (Ramot et al., 2016, 2017; Shibata et al., 2016, 2019). For example, using implicit neurofeedback, Shibata et al. (2016) showed that preferences for faces can be manipulated unconsciously with a high degree of specificity (i.e. by modulating different patterns of neural activity representing opposite directions of facial preferences *within* a brain region) and other studies have demonstrated implicit neurofeedback can be used to unconsciously manipulate perceptual confidence (Cortese et al., 2016) and physiological threat responses toward animals and objects (Taschereau-Dumouchel et al., 2018a). Emerging neurofeedback capabilities will be able to target more complex neural systems (Bassett and Khambhati, 2017), albeit within the limits to which the brain is robust to perturbations introduced through neurofeedback. Implicit neurofeedback can thus have nontrivial effects on mental function, and researchers who wish to use implicit neurofeedback are faced with unique challenges in obtaining informed consent from participants.

Informed consent is the process by which a researcher or clinician gets permission from a participant or patient before conducting an intervention. The Belmont Report specifies informed consent as part of a fundamental ethical principle for human subjects' research. Most, if not all, institutions consider the Belmont Report an essential reference for the institutional review board (IRB). Informed consent can only be obtained if the participant is of sound cognitive capacity and is provided with sufficient information to understand the potential risks and benefits of participation. Of course, implicit neurofeedback complicates the underlying assumption that the participant has sufficient information to evaluate the potential risks and benefits.

Given that many research studies rely on a degree of deception or secrecy of purpose, there is some allowance for this in the informed consent process. Deceptive instructions or omissions of information during the consent process can be authorized if it is deemed that the value and importance of a study can justify it, if the results cannot be obtained in some other way. However, sufficient information must still be provided to the participant to allow them to assess the risks and benefits of participation. In a situation involving deception or omission, it is important to consider that information that is sufficient for most people may not be sufficient for everyone.

For example, in the egg salad study discussed before, if a participant recently became vegetarian and disliked beans and dairy products, eggs could be a critical source of protein and the vitamin B12 for them. Inducing an aversion to egg salad in this case could actually result in a nontrivial dietary limitation. In short, the more information that is withheld, and the more deception that is used, the more risk there is that some participants may be unable to accurately assess risk of participation.

It is important to note that the degree to which implicit neurofeedback involves withholding information from participants can be considered as a continuum. On one end of the continuum are studies where participants do not even know that they are receiving neurofeedback. For example, Ramot et al. (2016) led participants to believe that they were participating in a neuroimaging study of reward responses in the brain, so participants had no idea that the reward they were receiving was dependent on their brain activity. More often, in implicit neurofeedback studies, participants are made aware of the fact that mental function is being trained by neurofeedback, but they are given no information about what aspect of mental function is trained (Amano et al., 2016; Cortese et al., 2016; Shibata et al., 2011, 2016). In some cases, participants may be given information regarding multiple possible types of mental function that may be trained, but are blinded to which type of training they will receive. Note that this approach involves a similar level of deception/omission as so-called explicit neurofeedback protocols that involve blinded control groups. This highlights that the explicit/implicit distinction may be less relevant from the perspective of ethical considerations than the degree of withheld information which varies in a dimensional sense across all neurofeedback protocols from those with full disclosure to those where participants do not even know they are being trained.

The more information is withheld during the consent process, the more responsibility falls on researchers and clinicians to avoid unwanted outcomes (Nakazawa et al., 2016). Extreme caution should be taken to ensure that the potential for any adverse effect is kept at minimum. One way to enhance informed consent but keep the participant unaware of the precise nature of the intervention is to let the participant know that the experiment involves placebo-controlled groups or conditions. This is a common procedure in clinically oriented neurofeedback experiments (Taschereau-Dumouchel et al., 2018a; Young et al., 2017; Zweerings et al., 2019).

Standards and guidelines set by IRBs and professional associations will often require that researchers debrief their study participants after completing the study. In the research context, a debriefing following participation is often included that provides more information to the subject regarding their training and its purpose. In cases of double-blind experiments, debriefing takes place once blinding is no longer critical for scientific purposes. In a nonresearch context, the mental aspects trained, and the purpose of that training should also be disclosed.

4 Conclusion

The goal of neurofeedback is to train participants to control specific aspects of brain activity. This technology is still in its infancy but holds much promise. It has potential as a personalized intervention tool for mental disorders which can be tested in double-blind and placebo-controlled settings (Chiba et al., 2019; Taschereau-Dumouchel et al., 2018b; Zahn et al., 2019). The ability to make causal inferences allows researchers the capacity to better study the relationship between the brain and behavior. Neurofeedback is only limited by how effectively its brain activity targets can be trained and how causally related those targets are to the mental functions of interest.

With new opportunities come new ethical challenges. The potential of neurofeedback to mold mental function raises concerns regarding threats to cognitive liberty that may be amplified over other noninvasive neurotechnologies if neurofeedback training proves capable of more effectively targeting specific mental processes. As fMRI neurofeedback is translated to more portable, accessible modalities, concerns arise regarding its potential social impact. Evidence for the efficacy of implicit neurofeedback amplifies these risks as well as risks to individual participants. Furthermore, the risk-benefit ratio of different neurofeedback protocols is not well understood given the limited data currently available on the longevity and reversibility of effects. In short, fMRI neurofeedback involves complex ethical concerns that deserve attention from the researchers who are developing its capacities for scientific and clinical application, and that should be communicated clearly to participants and to the public more generally.

The aim of this chapter is not to provide guidelines for the ethical conduct of fMRI neurofeedback research, which in our opinion, should be developed with input from the broader community. However, we can propose a few simple steps that could be helpful in fostering positive development. These include the collection of more follow-up data, not only monitoring persistence of the intended effects on mental function, but also incorporating broad screening of adverse events in both basic science and clinical studies. Careful management of prospective participants' expectations before neurofeedback training is also important. Fortunately, this is generally well managed at present due to existing regulatory frameworks governing human subjects' research. Finally, greater communication is needed on ethical issues between researchers, regulators, and the public.

As acknowledged early on, this is not intended as a comprehensive review of all the relevant ethical issues. Ethical issues are at risk of being overlooked as researchers often fail to imagine the personal impacts on participants or the potentially far-reaching consequences of increased dissemination in the context of technological development. However, it is important that we anticipate and reflect upon these issues before they become problematic. We hope this will provide a starting point for fruitful discussions on the ethics of fMRI neurofeedback.

Acknowledgments

MH was supported by the National Institute of Mental Health (R01 MH100068 and R61 MH115110), and the American Endowment Foundation (project titled “A neurofeedback booster for emotion regulation therapy”).

References

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., Watanabe, T., 2016. Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. Curr. Biol. 26 (14), 1861–1866. https://doi.org/10.1016/j.cub.2016.05.014.

Bard, I., Gaskell, G., Allansdottir, A., da Cunha, R.V., Eduard, P., Hampel, J., et al., 2018. Bottom up ethics – neuroenhancement in education and employment. Neuroethics 11 (3), 309–322. https://doi.org/10.1007/s12152-018-9366-7.

Bassett, D.S., Khambhati, A.N., 2017. A network engineering perspective on probing and perturbing cognition with neurofeedback. Ann. N. Y. Acad. Sci. 1396, 126–143. https://doi.org/10.1111/nyas.13338.

Battleday, R.M., Brem, A.K., 2015. Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: a systematic review. Eur. Neuropsychopharmacol. 25 (11), 1865–1881. https://doi.org/10.1016/j.euroneuro.2015.07.028.

Boorse, C., 2011. Concepts of Health and Disease. Philosophy of Medicine. Vol. 16 Elsevier B.V., https://doi.org/10.1016/B978-0-444-51787-6.50002-7.

Brühl, A.B., d’Angelo, C., Sahakian, B.J., 2019. Neuroethical issues in cognitive enhancement: modafinil as the example of a workplace drug? Brain and Neurosci. Adv. 3. https://doi.org/10.1177/2398212818816018, 239821281881601.

Bublitz, J.-C., 2013. My Mind Is Mine!? Cognitive Liberty as a Legal Concept. Springer, pp. 233–264, https://doi.org/10.1007/978-94-007-6253-4_19.

Buyx, A., 2015. Smart drugs: ethical issues. In: Handbook of Neuroethics, pp. 1192–1204, https://doi.org/10.1007/978-94-007-4707-4.

Chiba, T., Kanazawa, T., Koizumi, A., Ide, K., Taschereau-Dumouchel, V., Boku, S., et al., 2019. Current status of neurofeedback for post-traumatic stress disorder: a systematic review and the possibility of decoded neurofeedback. Front. Hum. Neurosci. 13 (July), 1–13. https://doi.org/10.3389/fnhum.2019.00233.

Coates McCall, I., Lau, C., Minielly, N., Illes, J., 2019. Owning ethical innovation: claims about commercial wearable brain technologies. Neuron 102 (4), 728–731. https://doi.org/10.1016/j.neuron.2019.03.026.

Cohen-Kadosh, R.C., Levy, N., O’Shea, J., Shea, N., Savulescu, J., 2012. The neuroethics of non-invasive brain stimulation. Curr. Biol. 22 (4), R108–R111. https://doi.org/10.1016/j.cub.2012.01.013.

Cohen-Kadosh, K., Lisk, S., Lau, J.Y.F., 2016. The ethics of (neuro) feeding back to the developing brain. AJOB Neurosci. 7 (2), 132–133. https://doi.org/10.1080/21507740.2016.1189979.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., Lau, H., 2016. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat. Commun. 7, 1–18. https://doi.org/10.1038/ncomms13669.

Cortese, A., Amano, K., Koizumi, A., Lau, H., Kawato, M., 2017. Decoded fMRI neurofeedback can induce bidirectional confidence changes within single participants. NeuroImage 149 (February), 323–337. https://doi.org/10.1016/j.neuroimage.2017.01.069.

deBettencourt, M.T., Norman, K.A., 2016. Neuroscience: incepting associations. Curr. Biol. 26 (14), R673–R675. https://doi.org/10.1016/j.cub.2016.05.041.

DeBettencourt, M.T., Cohen, J.D., Lee, R.F., Norman, K.A., Turk-Browne, N.B., 2015. Closed-loop training of attention with real-time brain imaging. Nat. Neurosci. 18 (3), 470–478. https://doi.org/10.1038/nn.3940.

Doruk Camsari, D., Kirkovski, M., Croarkin, P.E., 2018. Therapeutic applications of noninvasive neuromodulation in children and adolescents. Psychiatr. Clin. N. Am. 41 (3), 465–477. https://doi.org/10.1016/j.psc.2018.05.003.

Erler, A., 2011. Does memory modification threaten our authenticity? Neuroethics 4 (3), 235–249. https://doi.org/10.1007/s12152-010-9090-4.

Geraerts, E., Bernstein, D.M., Merckelbach, H., Linders, C., Raymaekers, L., Loftus, E.F., 2008. Lasting false beliefs and their behavioral consequences. Psychol. Sci. 19 (8), 749–753. https://doi.org/10.1111/j.1467-9280.2008.02151.x.

Hameed, M.Q., Dhamne, S.C., Gersner, R., Kaye, H.L., Oberman, L.M., Pascual-Leone, A., Rotenberg, A., 2017. Transcranial magnetic and direct current stimulation in children. Curr. Neurol. Neurosci. Rep. 17 (2), 1–25. https://doi.org/10.1007/s11910-017-0719-0.

Horstkötter, D., 2016. The ethics of novel Neurotechnologies: focus on research ethics and on moral values. AJOB Neurosci. 7 (2), 123–125. https://doi.org/10.1080/21507740.2016.1172135.

Ienca, M., Andorno, R., 2017. Towards new human rights in the age of neuroscience and neurotechnology. Life Sci. Soc. Policy 13 (1), 5. https://doi.org/10.1186/s40504-017-0050-1.

Ilieva, I., Hook, C., Farah, M., 2015. Prescription stimulants' effects on healthy inhibitory control, working memory, and episodic memory: a meta-analysis Irena. J. Cogn. Neurosci., 1–21. https://doi.org/10.1162/jocn.

Keynan, J.N., Meir-Hasson, Y., Gilam, G., Cohen, A., Jackont, G., Kinreich, S., et al., 2016. Limbic activity modulation guided by functional magnetic resonance imaging–inspired electroencephalography improves implicit emotion regulation. Biol. Psychiatry 80 (6), 490–496. https://doi.org/10.1016/j.biopsych.2015.12.024.

Keynan, J.N., Cohen, A., Jackont, G., Green, N., Goldway, N., Davidov, A., et al., 2019. Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. Nat. Hum. Behav. 3 (1), 63–73. https://doi.org/10.1038/s41562-018-0484-3.

Kimmel, H.D., 1986. The myth and the symbol of biofeedback. Int. J. Psychophysiol. 3 (3), 211–218. https://doi.org/10.1016/0167-8760(86)90029-2.

Klein, E., Goering, S., Gagne, J., Shea, C.V., Franklin, R., Zorowitz, S., et al., 2016. Brain-computer interface-based control of closed-loop brain stimulation: attitudes and ethical considerations. Brain Comput. Interfaces 3 (3), 140–148. https://doi.org/10.1080/2326263X.2016.1207497.

Lavazza, A., 2018. Freedom of thought and mental integrity: the moral requirements for any neural prosthesis. Front. Neurosci. 12, 1–10. https://doi.org/10.3389/fnins.2018.00082.

Linden, D., 2014. Brain Control: Developments in Therapy and Implications for Society, first ed. Palgrave Macmillan, New York, https://doi.org/10.1057/9781137335333.

Loftus, E.F., 2005. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. Learn. Mem. 12, 361–366. https://doi.org/10.1101/lm.94705.

Maier, L.J., Ferris, J.A., Winstock, A.R., 2018. Pharmacological cognitive enhancement among non-ADHD individuals—a cross-sectional study in 15 countries. Int. J. Drug Policy 58 (May), 104–112. https://doi.org/10.1016/j.drugpo.2018.05.009.

Maslen, H., Earp, B.D., Kadosh, R.C., Savulescu, J., 2014. Brain stimulation for treatment and enhancement in children: an ethical analysis. Front. Hum. Neurosci. 8, 1–5. https://doi.org/10.3389/fnhum.2014.00953.

Medaglia, J.D., Yaden, D., Helion, C., Haslam, M., 2019. Moral attitudes and willingness to induce cognitive enhancement and repair with brain stimulation. Brain Stimul. 12 (1), 44–53. https://doi.org/10.1016/j.brs.2018.09.014.Moral.

Megumi, F., Yamashita, A., Kawato, M., Imamizu, H., 2015. Functional MRI neurofeedback training on connectivity between two regions induces long-lasting changes in intrinsic functional network. Front. Hum. Neurosci. 9, 160. https://doi.org/10.3389/fnhum.2015.00160.

Meir-Hasson, Y., Kinreich, S., Podlipsky, I., Hendler, T., Intrator, N., 2014. An EEG finger-print of fMRI deep regional activation. NeuroImage 102 (P1), 128–141. https://doi.org/10.1016/j.neuroimage.2013.11.004.

Meir-Hasson, Y., Keynan, J.N., Kinreich, S., Jackont, G., Cohen, A., Podlipsky-Klovatch, I., et al., 2016. One-class FMRI-inspired EEG model for self-regulation training. PLoS One 11 (5), 1–17. https://doi.org/10.1371/journal.pone.0154968.

Nakazawa, E., Yamamoto, K., Tachibana, K., Toda, S., Takimoto, Y., Akabayashi, A., 2016. Ethics of decoded neurofeedback in clinical research, treatment, and moral enhancement. AJOB Neurosci. 7 (2), 110–117. https://doi.org/10.1080/21507740.2016.1172134.

Paret, C., Goldway, N., Zich, C., Keynan, J.N., Hendler, T., Linden, D., Cohen Kadosh, K., 2019. Current progress in real-time functional magnetic resonance-based neurofeedback: methodological challenges and achievements. NeuroImage 202. https://doi.org/10.1016/j.neuroimage.2019.116107, 116107.

Partridge, B.J., Bell, S.K., Lucke, J.C., Yeates, S., Hall, W.D., 2011. Smart drugs "as common as coffee": media hype about neuroenhancement. PLoS One 6 (11). https://doi.org/10.1371/journal.pone.0028416.

President's Council of Bioethics, 2003. Beyond Therapy. Biotechnology and the Pursuit of Happiness. Dana Press, New York.

Ramot, M., Grossman, S., Friedman, D., Malach, R., 2016. Covert neurofeedback without awareness shapes cortical network spontaneous connectivity. Proc. Natl. Acad. Sci. U. S. A. 113 (17), E2413–E2420. https://doi.org/10.1073/pnas.1516857113.

Ramot, M., Kimmich, S., Gonzalez-Castillo, J., Roopchansingh, V., Popal, H., White, E., et al., 2017. Direct modulation of aberrant brain network connectivity through real-time NeuroFeedback. elife 6, 1–23. https://doi.org/10.7554/eLife.28974.

Rance, M., Walsh, C., Sukhodolsky, D.G., Pittman, B., Qiu, M., Kichuk, S.A., et al., 2018. Time course of clinical change following neurofeedback. NeuroImage 181 (May), 807–813. https://doi.org/10.1016/j.neuroimage.2018.05.001.

Robineau, F., Meskaldji, D.E., Koush, Y., Rieger, S.W., Mermoud, C., Morgenthaler, S., et al., 2017. Maintenance of voluntary self-regulation learned through real-time fMRI neurofeedback. Front. Hum. Neurosci. 11 (March), 1–8. https://doi.org/10.3389/fnhum.2017.00131.

Ruiz, S., Lee, S., Soekadar, S.R., Caria, A., Veit, R., Kircher, T., et al., 2013. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. Hum. Brain Mapp. https://doi.org/10.1002/hbm.21427.

Sadato, N., Morita, K., Kasai, K., Fukushi, T., Nakamura, K., Nakazawa, E., et al., 2019. Neuroethical issues of the brain/MINDS project of Japan. Neuron 101 (3), 385–389. https://doi.org/10.1016/j.neuron.2019.01.006.

Schermer, M., 2015. Ethics of pharmacological mood enhancement. In: Handbook of Neuroethics, pp. 1178–1189, https://doi.org/10.1007/978-94-007-4707-4.

Sententia, W., 2006. Neuroethical considerations: cognitive liberty and converging technologies for improving human cognition. Ann. N. Y. Acad. Sci. 1013 (1), 221–228. https://doi.org/10.1196/annals.1305.014.

Sepulveda, P., Sitaram, R., Rana, M., Montalba, C., Tejos, C., Ruiz, S., 2016. How feedback, motor imagery, and reward influence brain self-regulation using real-time fMRI. Hum. Brain Mapp. https://doi.org/10.1002/hbm.23228.

Shibata, K., Watanabe, T., Sasaki, Y., Kawato, M., 2011. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science 334 (6061), 1413–1415. https://doi.org/10.1126/science.1212003.

Shibata, K., Watanabe, T., Kawato, M., Sasaki, Y., 2016. Differential activation patterns in the same brain region led to opposite emotional states. PLoS Biol. 14 (9), 1–27. https://doi.org/10.1371/journal.pbio.1002546.

Shibata, K., Lisi, G., Cortese, A., Watanabe, T., Sasaki, Y., Kawato, M., 2019. Toward a comprehensive understanding of the neural mechanisms of decoded neurofeedback. NeuroImage 188 (July 2018), 539–556. https://doi.org/10.1016/j.neuroimage.2018.12.022.

Snowball, A., Tachtsidis, I., Popescu, T., Thompson, J., Delazer, M., Zamarian, L., et al., 2013. Long-term enhancement of brain function and cognition using cognitive training and brain stimulation. Curr. Biol. 23 (11), 987–992. https://doi.org/10.1016/j.cub.2013.04.045.

Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J.D., Kawato, M., Lau, H., 2018a. Towards an unconscious neural reinforcement intervention for common fears. Proc. Natl. Acad. Sci. 115 (13), 3470–3475. https://doi.org/10.1073/pnas.1721572115.

Taschereau-Dumouchel, V., Liu, K.Y., Lau, H., 2018b. Unconscious psychological treatments for physiological survival circuits. Curr. Opin. Behav. Sci. 24, 62–68. https://doi.org/10.1016/j.cobeha.2018.04.010.

Thibault, R.T., MacPherson, A., Lifshitz, M., Roth, R.R., Raz, A., 2018. Neurofeedback with fMRI: a critical systematic review. NeuroImage 172 (September 2017), 786–807. https://doi.org/10.1016/j.neuroimage.2017.12.071.

Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M., 2017. Advances in fMRI real-time neurofeedback. Trends Cogn. Sci. 21 (12), 997–1010. https://doi.org/10.1016/j.tics.2017.09.010.

Westbrook, A., van den Bosch, R., Määttä, J.I., Hofmans, L., Papadopetraki, D., Cools, R., Frank, M.J., 2020. Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. Science 367 (6484), 1362–1366. https://doi.org/10.1126/SCIENCE.AAZ5891.

Westphal, A., Chow, T., Ngoy, C., Zuo, X., Liao, V., Storozuk, L., et al., 2019. Anodal Transcranial direct current stimulation to the left rostrolateral prefrontal cortex selectively improves source memory retrieval. J. Cogn. Neurosci., 1–12. https://doi.org/10.1162/jocn.

Wexler, A., Thibault, R., 2019. Mind-Reading or misleading? Assessing direct-to-consumer electroencephalography (EEG) devices marketed for wellness and their ethical and regulatory implications. J. Cogn. Enhanc. 3 (1), 131–137. https://doi.org/10.1007/s41465-018-0091-2.

Young, K.D., Siegle, G.J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., et al., 2017. Randomized clinical trial of real-time fMRI amygdala neurofeedbackfor major depressive disorder: effects on symptoms and autobiographical memory recall. Am. J. Psychiatr. 174 (8), 748–755. https://doi.org/10.1176/appi.ajp.2017.16060637.

Yuste, R., Goering, S., Agüeray Arcas, B., Bi, G., Carmena, J.M., Carter, A., et al., 2017. Four ethical priorities for neurotechnologies and AI. Nature 551 (7679), 159–163. https://doi.org/10.1038/551159a.

Zahn, R., Weingartner, J.H., Basilio, R., Bado, P., Mattos, P., Sato, J.R., et al., 2019. Blame-rebalance fMRI neurofeedback in major depressive disorder: a randomised proof-of-concept trial. NeuroImage: Clinical 24 (August), 101992. https://doi.org/10.1016/j.nicl.2019.101992.

Zuk, P., Torgerson, L., Sierra-Mercado, D., Lazaro-Munoz, G., 2018. Neuroethics of neuromodulation: an update. Curr. Opin. Biomed. Eng. 8, 45–59. https://doi.org/10.1016/j.cobme.2018.10.003.

Zweerings, J., Hummel, B., Keller, M., Zvyagintsev, M., Schneider, F., Klasen, M., Mathiak, K., 2019. Neurofeedback of core language network nodes modulates connectivity with the default-mode network: a double-blind fMRI neurofeedback study on auditory verbal hallucinations. NeuroImage 189 (January), 533–542. https://doi.org/10.1016/j.neuroimage.2019.01.058.