More Harm Than Good: Neurotechnological Thought Apprehension in Forensic Psychiatry

Mackenzie Graham & Phoebe Friesen

To cite this article: Mackenzie Graham & Phoebe Friesen (2019) More Harm Than Good: Neurotechnological Thought Apprehension in Forensic Psychiatry, AJOB Neuroscience, 10:1, 17-19, DOI: 10.1080/21507740.2019.1595787

To link to this article: https://doi.org/10.1080/21507740.2019.1595787

Published online: 09 May 2019.

Submit your article to this journal

Article views: 4

View Crossmark data
More Harm Than Good: Neurotechnological Thought Apprehension in Forensic Psychiatry

Mackenzie Graham, University of Oxford
Phoebe Friesen, University of Oxford

Advances in neurotechnology over the last several decades have led to a great deal of excitement, as individuals thought to be in persistent vegetative states have been found to be able to communicate within brain scans and paralyzed individuals have had the opportunity to move robotic arms through the use of brain–computer interfaces (Moxon and Foffani 2015; Owen et al. 2006). In “Ethical Issues to Consider Before Introducing Neurotechnological Thought Apprehension in Psychiatry,” Gerben Meynen considers the potential benefits that such neurotechnology might bring to the field of psychiatry. Noting that psychiatrists are often left in the dark when patients choose to withhold information from them, and that this can interfere with both diagnosis and treatment, Meynen argues that “reliance on first-person accounts is a serious vulnerability of psychiatric evaluations” (Meynen 2019, 7). “Neurotechnological thought apprehension” (NTA), he suggests, could bridge this epistemic gap for psychiatrists, both in clinical and forensic contexts. For example, NTA might allow a psychiatrist to determine whether a patient is withholding a past trauma, or whether a convicted individual might be eligible for an insanity plea.

While Meynen discusses several ethical issues that are likely to arise in such contexts, he tends to underestimate the inherent limitations of NTA. Recognizing such limitations, we argue, is critical to ensuring the ethical application of this technology. Meynen acknowledges this somewhat, suggesting that the criterion for using NTA in psychiatry should be “its added value compared to other techniques” (Meynen 2019). While he does not define such added value, one might assume that there is “added value” in NTA insofar as its use exceeds a minimum threshold of potential benefit and its costs do not surpass this benefit. In the forensic setting, a technology is beneficial insofar as it reliably contributes to convicting all and only those who are guilty of crimes, while costs will depend on the burdens of providing it, the resources required to obtain it, and the potential for a technology to frustrate the ends of justice (e.g., by generating false positives or false negatives). Accordingly, a highly burdensome and resource intensive means of obtaining evidence would only be justified if the evidence generated was highly reliable, but this would still have to be weighed against the costs of subjecting potentially innocent suspects to it.

In line with this, one might consider DNA evidence to be a good example of “added value.” Obtaining a DNA sample through a cheek swab is minimally burdensome, is not resource intensive, and the evidence generated by the sample is highly reliable. Accordingly, we can conclude that, in general, the potential benefits of DNA evidence are high, and the potential costs are low. Indeed, this balance of potential benefits over costs justifies the use of coercion to obtain a DNA sample in some cases, even if the individual providing the sample turns out not to be guilty of a crime. Does NTA meet this standard of minimal burden, proportional resource expenditure, and high reliability? There is little question that the use of NTA is more burdensome and requires more resources than DNA testing. For example, although they are fairly safe, functional magnetic resonance imaging (fMRI) scans are time-consuming, expensive, uncomfortable, and may be highly stressful (e.g., for individuals experiencing paranoia or psychosis, when NTA is framed as “mind reading”). However, even if technologies improve and infrastructures are put in place that make the use of NTAs affordable and accessible, significant worries related to the reliability of NTA are likely to remain. Three features that jeopardize the reliability of evidence based on NTA relate to complexity, intentionality, and interpretation; we consider each in the following.

First and foremost, the kind of mental states that would be useful for forensic psychiatry are highly complex, and likely too complex to be reliably interrogated by NTA. Meynen suggests that it would be useful for a psychiatrist to be able to verify whether an individual was in fact hearing voices at the time a crime was committed, but it is extremely unlikely that mental states of this kind could be localized to an area of the brain with such specificity that we could infer its content strictly based on the neural activation observed (Petersen and Sporns 2015). Rather, it is far more likely that such complex mental states will involve activation in a diverse network of...
brain regions. The mental imagery task pioneered by Owen and colleagues, which functions as a “proof of principle” for many of Meynen’s speculative applications of NTA, distinguishes between a participant imagining playing tennis or navigating their house. Imagining these tasks generates significant activation in spatially discrete areas of the brain: the supplementary motor area in the case of tennis, and the parahippocampal gyrus in the case of spatial navigation (Owen et al. 2006). In this case, the mental states of the participants can be plausibly inferred because the participants were instructed to imagine specific tasks, and activity in the expected brain areas was observed. The observed activation alone provides virtually no information about the specific content of the mental state, never mind its veridicality.

Compare this with more recent functional neuroimaging research, which found that healthy subjects demonstrate a highly similar pattern of neural activation when watching a suspenseful Alfred Hitchcock film (Naci et al. 2014). These results suggest that each participant was having a similar conscious experience of suspense in response to the stimulus. However, this is not the same as determining precise details of a person’s mental state, the kind of details that Meynen suggests would be useful to forensic psychiatry. Broadly similar conscious experiences (e.g., suspense) can elicit similar patterns of activation across individuals, because the same underlying cognitive processes are implicated in each, even though their content is very different. Yet individual brain areas often serve multiple purposes, and may even be associated with conflicting emotions (Phan et al. 2002). Thus, mental states that would be widely divergent in terms of their content—feeling suspense from a character about to be killed, compared to feeling suspense from two characters about to fall in love—could be indistinguishable via functional neuroimaging. And highly specific mental states, such as the belief that one’s friend is an alien (to borrow an example from Meynen), would require highly specific—indeed, nearly infinitely specific—patterns of activation in order to be identifiable.

A second barrier to the application of NTA in a forensic context is the central role of intention in such technologies. The examples of NTA discussed by Meynen rely heavily on the intentions of participants involved (e.g., intentionally imagining playing tennis) (Meynen 2019). If such intentions are not engaged, it isn’t clear that NTA could be performed on an unwilling subject. While Meynen briefly touches on this issue, he suggests that the susceptibility of fMRI to movement is the main difficulty. We think the problem is deeper than this. While subjects could easily distort the neural signal by moving in the scanner, there are also no real means of compelling a subject to adhere to the instructions of a mental task.

Finally, a significant degree of interpretation is required to draw conclusions about data drawn from NTA. The data acquired from functional neuroimaging may contain significant “noise” from other sources, which must be filtered out before the signal can be interpreted. An fMRI scan measures slices of brain activity at different time points, which must be corrected to bring all the slices into a single timepoint. Head motion must also be corrected, and the functional images must be aligned with the structural image of the individual. Each of these methods of “data smoothing” can influence the reliability of the data. For example, Comte and colleagues demonstrated that between 4.2% and 33% of responses to the mental imagery task were misclassified in a sample of healthy participants, depending on the algorithms used to interpret the data (Comte et al. 2015). The idea of ‘real time NTA,’ put forward by Meynen, is inconsistent with how functional neuroimaging data are generated. Neuroimaging data are not a direct representation of what is occurring in the brain, and thus, do not provide direct insight into a subject’s mental states.

Taken together, these features suggest that the applications Meynen envisages for NTA are unlikely to be within reach anytime soon. We should resist the temptation to introduce such technologies into the courtroom, heeding lessons from the cycle of hype and disappointment that have followed the introduction of many lie detection technologies in the past (Farah et al. 2014). Beyond the courtroom, we should be even more cautious in introducing NTA as a component of the psychiatrist’s toolbox. Important differences between clinical and forensic settings ought to influence how we evaluate the costs and benefits of using this neurotechnology. While forcing an individual to undergo a functional neuroimaging scan against their will in a legal context is a harm to them, this harm might be outweighed by the potential benefits of correctly convicting a violent criminal. Conversely, in a clinical context, it is unclear what might be gained from extracting information from a patient who is unwilling or unable to share it, while the costs in terms of trust, coercion, and damage to the patient-physician relationship are likely to be great. More broadly, we suggest the need for caution when speculating about the future advancements of any technology. The failure of a technology to meet even unrealistic expectations can lead to accusations of “hype,” which can undermine trust in the science and mask legitimate gains. The recent advances in functional neuroimaging are impressive enough, and present their own ethical challenges worthy of attention. Looking past them is likely to do more harm than good.

FUNDING

MG’s research is supported by a Wellcome Centre Grant to The Wellcome Centre for Ethics and Humanities (203132/Z/16/Z). PF is funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, grant BRC-1215-20008.
Thought Apprehension: The “True” Self and The Risks of Mind Reading

Andrea Lavazza, Centro Universitario Internazionale

In his article on the ethical issues related to neurotechnological thought apprehension (NTA) in psychiatry, Meynen (2019) highlights relevant concerns and provides a helpful framework to assess the use of such techniques. However, it is important to add a conceptual clarification that has significant consequences on the application of the NTA and on the permissibility of its use. The point is the following: With a sophisticated use of functional magnetic resonance imaging (fMRI) (or other means of brain imaging), what is really being “read” in our brains? We can currently focus on some cerebral areas and on the brain correlates of specific cognitive processes, but we cannot get a complete “picture” of the ongoing brain activity.

We can hypothesize that neurotechnological knowledge and tools will improve in the future and that it will soon be possible to have a finer understanding, at a molecular level, of what happens in our brain moment by moment. However, it does not seem that this should exempt us from addressing some interpretive questions that have important consequences also from an ethical point of view. First, it is necessary to distinguish between the idea of real-time monitoring on one hand, and on the other an analysis of the cerebral activations at the time when the examination is performed, in order to draw from the analysis of those activations a relatively predictive personal and behavioral profile.

In the first case, if it were really technically possible to monitor brain activity in real time, NTA techniques could tell us with a fair degree of approximation and with a minimal temporal advance what the agent is about to decide to say or do, often (perhaps in most cases) even before they are aware of it themselves. This would be a highly invasive modality, difficult to recommend from an ethical point of view even in relation to crime prevention, and not particularly useful for other purposes. In the second case, instead, the aim would be to infer from the instrumentally detectable brain activations the inclinations and tendencies of the analyzed subject. In this case, current knowledge and foreseeable technical progress do not seem to put us in a particularly favorable position, if the goal of NTA is to understand how the agent will tend to behave in relation to specific areas of conduct. As shown in the following, in fact, the environmental influence on the one hand and the conscious control on the other seem to reduce the area in relation to which the NTA can tell us something significant.

THE CASE OF RACIAL BIAS

Specific psychological tests have shown that many people have an implicit racial bias (Greenwald et al. 2009a). Unless one wants to embrace a dualistic perspective on the relationship between mind and brain, one can assume that implicit attitudes have a cerebral correlation that is instrumentally detectable as a specific neuronal activation, for example, by fMRI. Therefore, one could use the case of the implicit racist attitudes as a test for the conceptual problems that must be considered in relation to the use of the NTA. Obviously, it is undeniable that having implicit racist attitudes does not mean that

Address correspondence to Andrea Lavazza Centro Universitario Internazionale, Via Garbasso, Arezzo, 42-52100, Italy. E-mail: lavazza67@gmail.com