

How do We Conduct Fruitful Ethical Analysis of Speculative Neurotechnologies?

Lucie White

This is a preprint of a guest editorial that appeared in

***AJOB Neuroscience* (2019), 10(1): 1-4.**

The published version is freely accessible at

<https://doi.org/10.1080/21507740.2019.1599082>

Gerben Meynen (2019) invites us to consider the potential ethical implications of what he refers to as “thought apprehension” technology for psychiatric practice, that is, technologies that involve recording brain activity, and using this to infer what people are thinking (or intending, desiring, feeling, etc.). His article is wide-ranging, covering several different ethical principles, various situations psychiatrists might encounter in therapeutic, legal, and correctional contexts, and a range of potential incarnations of this technology, some more speculative than others. The speculative nature of the technologies under consideration raises a broader question about how to provide a fruitful ethical analysis of not-yet-realized neurotechnologies. Such analysis is certainly a valuable part of the ethics of neuroscience—neurotechnological developments can raise pressing and sometimes quite novel ethical issues, which should be carefully analyzed and anticipated before these technological developments become a reality. Or sometimes an identification of parallels with existing problems can allow us to delve into existing theoretical resources, giving us insight on how to best equip ourselves to handle problems before or as they arise.

Although Meynen’s article raises some highly relevant potential implications of some “thought apprehension”-based neurotechnologies, which are certainly worthy of further exploration, I suggest that Meynen’s analysis is beset by two problems. First, some of the technologies that form the focus of his discussion are too speculative, and second, due to the many issues under consideration, some problems are not adequately characterized, relevant ethical principles are misidentified, and valuable literature is not sufficiently engaged with. Consideration of these weaknesses might allow us to better see how to conduct a fruitful analysis of speculative neurotechnologies.

How speculative is too speculative?

Meynen correctly notes that it is important for ethicists to anticipate problems with future technological developments, prior to a point where they could be introduced into (psychiatric) practice. There is, however, an important caveat to this contention. We need some indication of what kind of technological developments might be on the horizon, both because we should focus our attention on ethical issues we have reason to suspect we might encounter, and so we can identify and analyze potential ethical issues.

In his account of the current state of “thought apprehension” technology, Meynen cites studies in which participants are asked to visualize a specific concept or activity, which activates a certain pattern of brain activity, which can then be associated with an answer to a question. He also cites some studies in which some additional information can be gleaned from brain activity, such as the presence of auditory hallucinations, or suicidal ideation. Some of the potential future technologies that Meynen goes on to consider are directly related to these developments. He discusses, for example, a robotic arm, which could be controlled by a “locked-in” patient, using the same principle as the first set of studies—the pattern of brain activity accompanying a specific thought could be associated with a certain direction of movement (which, as Meynen points out, could raise interesting questions about moral responsibility).

However, some of the technologies that Meynen considers are far more removed from current developments. The most speculative technology Meynen considers is “live” mind-reading, which he imagines could ascertain all thoughts that occur to a person within a certain period of time. The leap from current technological developments to this technology faces serious methodological hurdles. Current techniques proceed by identifying a pattern of brain activity that is associated with a particular thought, which “has to be painstakingly established by getting a person to think a particular thought, while their concurrent brain activity is measured” (Haynes 2012, 33). It might be possible to use basic thoughts as building blocks combined to establish other thoughts that have not been directly measured (Haynes 2012), but it also might be the case that we cannot identify any constituent parts of thoughts that allow us to infer any other thoughts (Bayne 2012). If this is the case, such a technology can never eventuate. Even if we can make these inferences, there are still significant barriers to implementation. Certain types of thoughts correlate with significantly different brain activity between different people, making it impossible to train an algorithm on data from another person. The plasticity of the brain also presents a consequential problem— “continuous learning and change of connotations” (Haynes 2012, 33) could affect the relationship between a single person’s brain activation patterns and thoughts over time. These fundamental methodological challenges suggest that live mind-reading

is likely to “remain fictional for the foreseeable future” (Haynes 2012, 32), and may never be realized.

Such considerations should lead us to question the value of speculation about the ethical implications of introducing such a technology (see Jones, Whitaker and King 2011). There is also, however, a second problem with this speculation—even if we think such an endeavor is worthwhile, and even if we think the technology might eventuate, we do not know enough about what this technology might look like to identify pertinent ethical issues. Meynen imagines, for example, live mind-reading being used in place of a psychiatric assessment of a defendant. One potential ethical issue envisioned here is that “all kinds of mental content may be collected based on brain activity, and it may be very difficult to interpret their [sic] quality and significance. For example, suppose that brain activity related to sexual aggression towards others is registered: Is that just a fantasy, or an inclination, or an actual intention?” (2019, 8).¹ It might be the case that the brain activity for an intention and a fantasy could be difficult to distinguish. But Meynen provides no evidence to support this contention. It might equally be the case that the brain activation patterns correlating with these mental events are quite different. Thinnes-Elker and colleagues, for example, survey studies that identify specific brain activity associated with intention (2012). The point here is that we don’t, at this stage, know enough about what this potential technology would look like to know whether this will be an issue or not. It could be that even if this technology ever eventuates, it will bear little resemblance to what Meynen pictures (see Jones, Whitaker and King 2011). We are not yet in a position to adequately evaluate the ethical implications of this technology.

Fruitful ethical analysis

Let’s return to the example of live mind-reading, which Meynen discusses under the ethical principle of confidentiality. After briefly noting the largely familiar problems over the question of when to break confidentiality in therapeutic contexts, Meynen focuses on mind-reading technology in forensic contexts, or, more specifically, where psychiatrists are called upon to conduct evaluations or provide opinions for use in legal proceedings. He notes that the obligation of confidentiality is limited here—the purpose of the psychiatrist’s evaluation is to reveal pertinent information to the court (and perhaps indirectly to the media and the public).

¹ It should be noted that it’s not clear what Meynen means by intention here, as he seems to be talking about a defendant who has potentially already committed a crime – a past intention of the act in question? A current intention of a different, possibly related act? The former raises additional difficulties for the picture Meynen has in mind here – must the thoughts occur consciously to the person during the period of time in which he is monitored? Or might brain activity be used to infer other psychological elements? It’s not yet clear what this could possibly look like, if something along these lines is at all possible.

Putting the already-mentioned problems about identifying the type of mental event experienced aside, Meynen imagines a scenario in which thought-apprehension technology reveals the presence of sexual fantasies in a man charged with violence against his neighbor. The crucial question in determining whether this information should be shared by the psychiatrist, Meynen plausibly contends, is whether this information is relevant to the matter at hand. “But what counts as relevant? ... The answer might depend on the nature of those fantasies, but what exactly?” (2019, 8).

After raising this preliminary question, Meynen moves on to the next issue. But a glance at existing ethical literature on the issue of determining the relevance of material gleaned from psychiatric interviews for court proceedings reveals that this is not a novel problem. Such information is sometimes revealed in a traditional interview setting. Because of this, there are well-established legal and ethical standards that provide guidance concerning how to determine whether information is relevant in current forensic psychological practice (Glancy et al. 2015; Appelbaum 1984), with a specific focus on which information it is appropriate for the psychiatrist to convey, and to whom (Appelbaum 1990). Information about a defendant’s “fantasies, sexual preferences, and family relations” (Appelbaum 1984) is often excluded from reports, and the circumstances under which this kind of information can be invoked are carefully constrained (Glancy et al. 2015). Might this existing literature provide adequate or helpful guidance in dealing with this new technology? Or would the introduction of this new technology pose new problems that can’t be dealt with by existing ethical standards? Meynen gives us no indication of this. I suggest, putting the speculative nature of the technology aside, that there are three problems with this analysis. First, the problem is not adequately characterized—is this a novel problem, does it present new challenges? Second, highly relevant literature on what seems to be the same problem does not receive consideration, diverting attention away from a possible means of tackling this problem. This second issue might stem from a third—in couching this forensic discussion of relevant evidence under a therapeutic principle of confidentiality, the connections to ethical literature on relevance in forensic psychiatry may not be immediately apparent. We can see other examples of each of these issues in the text.

Characterization of the problem

In his discussion of competence, Meynen considers whether patients in a presumed vegetative state might be able to communicate treatment decisions via functional magnetic resonance imaging (fMRI). With reference to Jim Drane’s “sliding scale” of competence, he suggests that such patients might be able to make low-stakes decisions, but not high-stakes decisions. But what

criteria for competence is Meynen using here, and why do we have reason to think they would not be met at the high end of the scale? The problem here might be with the reliability of the technology in successfully communicating the patient's wishes, but this is belied by Meynen's suggestion that there would be no problem at the low end of the scale. Might there specifically be problems with establishing Drane's (1985) high-stakes criteria of appreciation and rational decision making? Are we relying on a fictional scenario in which the patient can in principle communicate in a sophisticated manner, or are there limitations on what can be effectively communicated? With no indication of the answers to any of these questions, it is difficult to make sense of the problem.

Engagement with relevant literature

In Meynen's discussion of coercion, he considers using information gleaned from functional magnetic resonance imaging (fMRI) to release from a psychiatric facility an offender who would otherwise not be regarded as a candidate for release. He suggests that although this offer can only improve the offender's situation, it may pose problems because it might subjectively be regarded as coercive: "It seems not unlikely that the patient may perceive the offer as a threat" (2019, 10). He cites an article by George Szumukler and Paul Appelbaum in making these claims, but although Szumukler and Appelbaum argue that the patient's subjective perception "may indicate issues that need to be explored and clarified with the patient" (2008, 237), they advocate an "objective" standard for determining what counts as coercion, with reference to a "moral baseline" (2008, 236)—will the recipient be made worse off, or is he being denied some right? Because there is no right to early release from prison, this offer, for Szumukler and Appelbaum, would not count as a coercive threat. Bioethicists Ruth Faden and Tom Beauchamp (who are not mentioned by Meynen) do advocate for subjective elements in determining whether an offer is coercive, based on whether the offer is easily resistible or welcomed (Faden and Beauchamp 1986). However, because it is likely that an offer of early release from prison will be welcomed, this does not seem to support Meynen's argument either.² Engagement with the vast literature on coercion in psychiatric and medical contexts could help with the analysis of this issue, and in determining whether this offer should be seen as coercive and why.

² Meynen might contend, against this, that the offer is not likely to be welcomed because it is invasive – he compares it to deep brain stimulation (which I have argued, in a different context, is not likely to be welcomed and should thus not be offered in exchange for release from prison (Hübner and White 2016)). This however, is an unconvincing analogy – DBS involves surgical intervention in the brain, which involves, among other things, risk of infection, haemorrhage (De Ridder et al. 2009) mortality, and disabling morbidity (Canavero 2014). More needs to be said about why this procedure should be seen as similarly invasive.

Identification of pertinent principles

Meynen's discussion of trust draws out a problem that was only possible and implicit in the initial example—it approaches a forensic issue with a therapeutic principle. It is unclear whether he thinks that “the central role trust plays in medicine—and therefore also in psychiatry” (2019, 9) applies to forensic contexts too. But he expresses concern that neurotechnological lie detection might undermine a trust relationship in forensic contexts by expressing “the psychiatrist's distrust” (9) of the evaluatee. However, the potential problem with this hinges on the assumption that the two-way trust relationship that Meynen emphasizes as so important in therapeutic contexts—the necessity of the doctor exhibiting trust in the patient, so the patient can trust the doctor—has any application in this setting. This misses a central issue in forensic contexts—a primary concern here is rather that the evaluatee will be too trusting of the psychiatrist (Glancy et al. 2015; Appelbaum 1984)—it is important that she understands that the psychiatrist, in this context, cannot be relied on to uphold therapeutic standards of keeping her confidence and putting her interests first. In order to avoid mischaracterizing the ethical issues at stake here, it is important to recognize the “entirely different ethical framework” (Appelbaum 1997, 445) in which forensic psychiatrists work, and to avoid applying principles from one context to another without adequate consideration.

Conclusion

An examination of Meynen's article can help us to think critically about how to conduct fruitful ethical analysis of speculative neurotechnologies. I have suggested that Meynen's consideration of the ethical issues stemming from neurotechnological thought apprehension contains two types of problems: First, some of the technologies he considers are too speculative, which means both that this might not be a valuable use of ethical resources, and that we do not know enough about the potential technology to identify the ethical implications. Second, largely due to the broad nature of his inquiry, he has not gone into enough detail to adequately characterize some problems, to draw attention to and engage with existing, relevant debates, and to identify the specific ethical principles applicable to a particular psychiatric context.³ It might be maintained that this is not the point of such an article—it just aims to draw preliminary attention to some potentially relevant ethical issues. But these features risk mischaracterizing or misidentifying the problems that may arise, and drawing attention away from resources that might help to approach them. I suggest that a more sustained analysis of a particular ethical issue that we are in a

³ This is certainly not to say that Meynen's article does not make a valuable contribution – he raises some ethical issues of great import that are worthy of further, sustained exploration, particularly in his discussion of responsibility.

position to identify and characterize, with a nuanced appreciation of the potential relevance of existing work and existing ethical concepts and principles, might be a better way to engage in ethical analysis of speculative neurotechnological developments.

References

AAPL practice guideline for the forensic assessment. 2015. *Journal of the American Academy of Psychiatry and the Law*, 43(2 Suppl), S3-53.

Appelbaum, P. 1984. Confidentiality in the forensic evaluation. *International Journal of Law and Psychiatry*, 7(3-4), 285-300.

Appelbaum, P. 1997. Ethics in evolution: the incompatibility of clinical and forensic functions. *American Journal of Psychiatry*. 154(4), 445-46.

Appelbaum, P. 1990. The parable of the forensic psychiatrist: ethics and the problem of doing harm. *International Journal of Law and Psychiatry*, 13(4), 249-59.

Bayne, T. 2012. How to read minds. In *I know what you're thinking: Brain imaging and mental privacy*, ed. S. Richmond, G. Rees, and S. Edwards, 41-58. Oxford: Oxford University Press.

Canavero, S. 2014. Criminal minds: Neuromodulation of the psychopathic brain. *Frontiers of Human Neuroscience* 8(124), 1-3. Available at: <http://dx.doi.org/10.3389/fnhum.2014.00124>

De Ridder, D., B. Langguth, M. Plazier, and T. Menovsky. 2009. Moral dysfunction: Theoretical model and potential neurosurgical treatments. In *The moral brain: Essays on the evolutionary and neuroscientific aspects of morality*, ed. J. Verplaetse, S. Vanneste, J. Braeckman, and J. Schrijver, 155–84. Dordrecht: Springer Netherlands.

Drane, J. 1985. The many faces of competency. *Hastings Center Report*, 15(2), 17-21.

Faden, R., and T. Beauchamp. 1986. *A history and theory of informed consent*. New York: Oxford University Press.

Haynes, J. D. 2012. Brain reading. In *I know what you're thinking: Brain imaging and mental privacy*, ed. S. Richmond, G. Rees, and S. Edwards, 29-40. Oxford: Oxford University Press.

Hübner, D. and L. White, Neurosurgery for psychopaths? An ethical analysis. *AJOB Neuroscience*, 7(3), 140-149.

Jones, G., M. Whitaker and M. King. 2011. Speculative ethics: Valid enterprise or tragic cul-de-sac? In *Bioethics in the 21st Century*, ed. A. Rudnick, IntechOpen, Available at: <https://doi.org/10.5772/19684>

Meynen, G. 2019. Ethical issues to consider before introducing neurotechnological thought apprehension in psychiatry. *AJOB Neuroscience*, 10(1), 5-14.

Szmukler, G., and P. Appelbaum. 2008. Treatment pressures, leverage, coercion, and compulsion in mental health care. *Journal of mental health*, 17(3), 233-244.

Thinnes-Elker, F., O. Iljina, J. K. Apostolides, F. Kraemer, A. Schulze-Bonhage, A. Aertsen, and T. Ball. 2012. Intention concepts and brain-machine interfacing. *Frontiers in Psychology*, 3(455), 1-10. Available at: <https://doi.org/10.3389/fpsyg.2012.00455>