



## Ethical Issues to Consider Before Introducing Neurotechnological Thought Apprehension in Psychiatry

Gerben Meynen

To cite this article: Gerben Meynen (2019) Ethical Issues to Consider Before Introducing Neurotechnological Thought Apprehension in Psychiatry, AJOB Neuroscience, 10:1, 5-14, DOI: [10.1080/21507740.2019.1595772](https://doi.org/10.1080/21507740.2019.1595772)

To link to this article: <https://doi.org/10.1080/21507740.2019.1595772>



© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 09 May 2019.



[Submit your article to this journal](#)



Article views: 342



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)

# Ethical Issues to Consider Before Introducing Neurotechnological Thought Apprehension in Psychiatry

Gerben Meynen, VU University Amsterdam and Utrecht University

When it becomes available, neuroscience-based apprehension of subjective thoughts is bound to have a profound impact on several areas of society. One of these areas is medicine. In principle, medical specialties that are primarily concerned with mind and brain are most likely to apply neurotechnological thought apprehension (NTA) techniques. Psychiatry is such a specialty, and the relevance of NTA developments for psychiatry has been recognized. In this article, I discuss ethical issues regarding the use of NTA techniques in psychiatric contexts. First, I consider the notion of neurotechnological “thought apprehension,” as well as some limitations of present-day NTA applications. Next, I identify ethical priorities for its possible future use in psychiatry. The topics I explore concern key (bio)ethical issues: confidentiality, trust and distrust, consent and coercion, and, finally, responsibility. I conclude that mental health-related use of NTA entails some specific ethical concerns that deserve careful attention before introducing these technologies in psychiatric practice.

**Keywords:** Neuroethics, neurolaw, mind reading, psychiatry, neurotechnology

## INTRODUCTION

Neurotechnological apprehension of subjective thoughts, when it becomes available, is bound to have a profound impact on several areas of society. One of these is medicine. In principle, medical specialties that are primarily concerned with the mind and brain are most likely to apply these neurotechnological thought apprehension (NTA) techniques. Psychiatry is such a specialty.<sup>1</sup> In this article, I consider ethical challenges regarding future psychiatric use of NTA, focusing primarily on forensic psychiatric practice. Although NTA is sometimes hyped in the media, it is important to note that at present, it is possible only on a very limited scale, basically in research settings. Yet because of its potentially far-reaching implications, ethical qualms have already been raised about (near-future) use of NTA.<sup>2</sup> Notably, normative

reflection should be in advance<sup>3</sup> of such developments before they are introduced in professional—including biomedical—practices.

However, before considering ethical issues, I briefly look at the concept of NTA and at the current state and limitations of NTA applications. In addition, I discuss the possible relevance of the applications for the field of psychiatry. I then explore several ethical challenges and priorities for NTA in psychiatry, especially involving confidentiality, trust and distrust, consent and coercion, and responsibility. Finally, since it is always unethical to use inappropriate techniques, I pay attention to concerns regarding the ecological validity of NTA in the specific context of psychiatry. The ethical challenges, I argue, have to be well considered before NTA is introduced in psychiatric practices.

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

1. The relevance of NTA regarding psychiatry is reflected in the literature: Richmond et al. (2012), Linden (2012), Fuchs (2006) and more specifically regarding forensic psychiatry Meynen (2017); Silva (2009).

2. See Shen (2013), Ienca and Andorno (2017), Roskies (2015), and references throughout this article.

3. See Nadelhoffer and Sinnott-Armstrong (2012) on neurotechnologies in the courtroom.

Address correspondence to Gerben Meynen, Philosophy, VU University Amsterdam, De Boelelaan 1105, Amsterdam, 1081HV, Netherlands. E-mail: [g.meynen@vu.nl](mailto:g.meynen@vu.nl)

## NEUROTECHNOLOGICAL THOUGHT APPREHENSION

### NTA's Current Limitations and Its Relevance for Psychiatry

The term neurotechnological thought apprehension, as used in this article, refers to neuroscience-based techniques that “detect” mental states or subjective phenomena. In the literature, alternative terms are neuroscientific “mind reading” or “brain reading.”<sup>4</sup> In fact, even though the term “thought” is used, NTA is meant to cover techniques that—more broadly—apprehend mental states, such as thoughts, intentions, inclinations, biases, desires, feelings, or levels of consciousness. Still, the main focus of this article is on apprehending a person's thoughts. On a technical level, the procedure has to involve neurotechniques (e.g., functional magnetic resonance imaging [fMRI] or electroencephalography [EEG]) but it may be combined with other technologies or methodologies as well, such as machine learning algorithms. On a theoretical level, NTA as discussed in this article neither implies nor requires a commitment to a specific conception of the mind–brain relationship (such as physical reductionism). In this sense, NTA is not dogmatic but pragmatic: As long as neuroimaging of some sort provides helpful (additional) information about a person's thoughts, intentions, and so on, it can be considered an NTA technique.

Using neurotechnologies, then, how successful is “thought apprehension” at this moment? It is fair to say that, in general, NTA is possible to some limited extent in research settings. An example is the often-cited study reporting that fMRI scanning made it possible to communicate with a patient in a vegetative state (Monti et al. 2010). The patient was instructed to answer questions by thinking about playing tennis or navigating his house—meaning either “yes” or “no”—which resulted in different brain activation patterns that could be detected by fMRI. This is a remarkable result, which has led to much neuroethical debate about (future) applications of such techniques. Yet, notably, this was a research setting, and even today, several years later, this type of mind reading is still not possible in clinical practice. Another more recent example, however, involved “reading” the physics concepts subjects were thinking about (Mason and Just 2016). These subjects, who had a background in physics or engineering, were instructed to think about abstract concepts such as gravity and velocity. Using fMRI combined with machine learning, it was possible to detect which concepts the subjects were contemplating. The study, however, involved a multiple-choice design, in which both the subjects and the computer could only “choose” between predetermined options,

which is a clear limitation. Therefore, it is still far from real-time mind reading, but it is suggestive of such a possibility in the near future.

Arguably, the most robust results—that also come closest to clinical NTA application—involve brain–computer interfaces (BCIs) used in locked-in and paralyzed patients (Soekadar et al. 2015). The BCI may enable patients to communicate with others by registering brain activity and selecting letters/words on a screen. In other instances, a robotic arm may be controlled based on detected brain activity patterns. Whereas the latter type of application is robotic, it still contains an NTA component in which a patient's intentions are somehow “registered” by technology and lead to robotic action. So, even though the “apprehension” in itself is not the ultimate goal, it is an essential part of the procedure.

A line of study that is specifically relevant for psychiatry concerns the “reading” of brain activation related to auditory hallucinations, an important psychopathological phenomenon. Psychotic patients may hear one or more voices that talk to them or among themselves, providing sometimes constant commentary or commands. Using fMRI, it has been possible to relate brain activity to the occurrence of auditory hallucinations aiming to “capture” them (Leroy et al. 2017). Another study that is relevant from a psychiatric perspective involved detecting “neural representations” of concepts related to death and life. Using fMRI combined with machine learning algorithms, researchers were able to identify individuals with suicidal ideation. They could even distinguish between those who had attempted suicide and those who had not (Just et al. 2017).

Most telling is that we cannot at present read off a patient's precise thoughts from a scan without having an already determined template for interpreting them, such as the multiple-choice design study already described. In general, NTA results up till now amount to what can be considered “proof of principle,” and there are still considerable obstacles and limitations for practical use. Since this article is about ethical challenges regarding future NTA use in psychiatry, it is beyond its scope to further discuss in detail the various shortcomings of current techniques. However, I address some further issues in the final section of this article on ecological validity.

### Reliance on First-Person Accounts and NTA in Psychiatry

Current psychiatric assessment has to rely to a large extent on what the patient tells the mental health professional. Even though in many circumstances the person's words constitute a highly valuable source, there are limitations. For instance, a patient may choose not to share certain information. A traumatized patient may not reveal the traumatic event, because talking about it is too uncomfortable. A patient with obsessive-compulsive disorder may not share distressing sexual intrusions

4. This, with what follows about the concept of NTA, is in accordance with (Meynen 2018a), where the term “brain-based mind reading” is used. On (ethical issues concerning) using these techniques in forensic psychiatry, see also Meynen (2017; 2018b).

because she is too ashamed. A paranoid patient may not share relevant information because she doesn't trust the psychiatrist. Not sharing such information in these cases may well affect the diagnostic process and related treatment decision making. Thus, reliance on first-person accounts is a serious vulnerability of psychiatric evaluations.<sup>5</sup> Therefore, if "the possibilities of neurotechnological mind-reading that we have today allow access to mental states without 1st person overt external behavior or speech" (Evers and Sigman 2013, 895) this will certainly be of interest to psychiatry.

Often, a reliance on first-person accounts becomes a crucial and challenging element in forensic psychiatry—broadly defined as psychiatry at the intersection with the law. The reason is the (presumed) increased risk of lying, malingering, and selective information sharing by the examinee in forensic contexts.<sup>6</sup> Consider the psychiatric evaluation of a defendant.<sup>7</sup> If anyone, the defendant knows what he felt at the moment of the crime, what he thought, and what he aimed to do and why. A defendant may tell the psychiatrist that he was convinced that his friend was actually an alien from another planet who was on the verge of killing him and that, therefore, he defended himself by attacking his friend. If this account is truthful, it could well support the conclusion that the defendant was deluded at the time of the crime and, more precisely, that he delusionally believed he was acting in self-defense. All of this can be critical to the issue of legal insanity. But is the defendant's account to be believed and relied on? How is one to verify its truthfulness? In principle, NTA (such as brain-based lie detection) could be helpful here. If brain-based lie detection would show that the defendant is probably truthful, the reliability of his testimony may well increase. Clearly, an honest defendant may still be mistaken, or somehow have trouble giving an accurate account of the situation, but the chance that the account is reliable increases, as long as the NTA is reliable.

Another possible NTA application in psychiatry could involve voices commanding patients to perform criminal actions. For instance, a voice may have ordered a defendant to attack his neighbor. Suppose that during an evaluation of this defendant a forensic psychiatrist asks the question: "Do you hear voices?" The defendant, however does not respond, because the voice orders him not to speak with the psychiatrist. How, then, is one to establish the presence of a voice in this silent defendant? Suppose, however, that the fMRI's "capturing of auditory hallucinations" technique were available in forensic

psychiatric practice; then "detecting" the voices in this defendant could not only shed more light on the matter of legal insanity and on future dangerousness, but also on treatment strategies that might not only reduce risks of future violence but possibly also improve the person's quality of life.

In fact, NTA may be particularly relevant at the interface of psychiatry and the law, as other interests than those related to health, such as avoiding punishment, may be at stake. Therefore, people may be tempted to conceal the (whole) truth about their symptoms or just lie about them.

In sum, given psychiatry's profound interest in the mind and mental capacities, and, more specifically, given its sometimes challenging reliance on first person accounts, NTA could be of considerable value to it.<sup>8</sup> In the next sections, I explore ethical concerns for possible future use of NTA in psychiatric practice regarding confidentiality, (dis)trust, (coerced) consent, and responsibility.

## ETHICAL CONCERNS

### Confidentiality

In medicine, the duty of confidentiality extends back to Hippocrates.<sup>9</sup> The health care professional has the obligation to keep medical information confidential, which would also be true for information obtained by NTA. Because "apprehending" a person's thoughts could yield information about what goes on in the person's mind, NTA-derived information in psychiatry would have to be treated with utmost care, which may involve increased data security measures.

Alternatively, there may be situations in which otherwise confidential information needs to be judiciously disclosed. As Roberts writes in *A Clinical Guide to Psychiatric Ethics*:

There are circumstances in which the risks associated with nondisclosure of the patient information in mental health care are very great and outweigh the individual's privilege of confidentiality. Reporting clinical findings becomes ethically and legally justified when there is a serious and imminent threat of physical harm to an identifiable and specific person, when breaking the confidence is likely to do good and to prevent harm (e.g., protection of the intended victim), and when other efforts to address the

5. See also Linden (2012) and Meynen (2017, 2018b).

6. Malingering is (intentionally) faking signs and symptoms of an illness (or aggravating them) in order to achieve a certain goal (such as being considered legally insane). On malingering in (forensic) psychiatry see Rogers (2012) and Feuerstein et al. (2005).

7. See on this topic also Linden (2012).

8. To be clear, these points (and what follows) are made not to argue that NTA should be used in psychiatry, but to show that future use of NTA in psychiatry is not improbable, and therefore deserves consideration also from an ethical perspective.

9. Beauchamp and Childress (2013) and see Roberts (2016). Roberts (2016, 73) also writes: "This duty [of confidentiality] derives from the broader philosophical concept of privacy ...." On confidentiality and NTA in forensic psychiatry, see also Meynen (2017).

situation have failed or are insufficient clinically or legally.<sup>10</sup>

Health care professionals are not only justified in reporting these cases, they may well have the legal and ethical obligation to report and thus to breach confidentiality. Could NTA data provoke such a breach? While confidentiality dilemmas abound in medicine, NTA could add considerably to them because it could provide so much information about mental states—which might be interpreted by the psychiatrist as presenting an unacceptable degree of risk.

What further complicates confidentiality obligations is that even in those cases in which a health professional reports, it is crucial not to disclose too much information but “only what is absolutely necessary” (Roberts 2016). Also, “How much information is too much? How much information is too little?” (Roberts 2016, 89).

Forensic psychiatric settings only aggravate these dilemmas. Here, there is no “doctor–patient” relationship, but rather an “examiner–examinee” relationship (Gutheil 2005, 346). Arguably the most well-known context concerns an assessment of a defendant’s legal insanity, where the psychiatrist is not consulted by the defendant for cure or alleviation of certain symptoms, but rather performs an evaluation as requested by a third party: the court, the lawyer, or the prosecution. While psychiatrists normally should not reveal any information to others, forensic psychiatry might require the psychiatrist to reveal information to the court. That is why the assessment takes place. Everything that is observed by the psychiatrist may be included in the report or in the testimony before the court. And as far as the court procedures are open to the public, the public may hear about them as well. They could even be printed or broadcast by the media. Obviously, then, forensic psychiatric situations limit confidentiality in significant ways (Gutheil 2005, 348).

Furthermore, while a typical conversation in forensic psychiatry involves psychiatrists asking questions with defendants responding, NTA could just register all thoughts that come up in a particular period of time, including thoughts that are not “answers to the question.” Clearly, this is a long way off. Yet if it were to become a technical possibility, such “detected” thoughts might concern certain hopes (“I hope this will soon be over”) or fears (“The guy frightens me”), or fantasies. To which extent should NTA information about such thoughts be shared with third parties, like the judge and jury? One might argue that as far as the NTA information is relevant, it can be shared, which seems a reasonable answer. But what counts as relevant? Suppose NTA reveals certain sexual fantasies in a person charged with violence against his neighbor. To what extent would that

be “relevant” and sharable with the court (and perhaps indirectly with the public and media)? The answer might depend on the nature of those fantasies, but how exactly?

Of course, all of this supposes that the findings are accurately interpreted. But they might not be. Real-time NTA—live mind reading—may function like a “trawl”: All kinds of mental content may be collected based on brain activity, and it may be very difficult to interpret their quality and significance. For example, suppose that brain activity related to sexual aggression toward others is registered: Is that just a fantasy, or an inclination, or an actual intention? At present, such “catching” of free-floating thoughts is impossible, but using NTA it might become a possibility—and the interpretation may be very challenging. Yet it is precisely the psychiatrist’s interpretation that becomes crucial regarding whether confidentiality is breached or not.

Finally, even though NTA techniques may provide new information, the possibility that it could be incorrect needs to be recognized, not only in theory, but also in psychiatric practice. Almost certainly, there will be a margin of error, and specificity and sensitivity are unlikely to be perfect. The level of specificity of the test is important when considering the duty to breach confidentiality. In fact, the information provided by NTA will be probabilistic in nature.

These are new questions that psychiatrists may have to answer when they start using NTA in a forensic context. Of note, psychiatry is the medical specialty that has a deep interest in these mental states, implying that much of what might be detected could be relevant. The pressure this places on the psychiatrist’s professional discretion and judgment seems extraordinary, indeed, unprecedented.

### Trust and Distrust

Trust is one of the core principles in medicine. It is the basis for a working alliance, and related to other principles such as confidentiality and beneficence. According to Hall, “trust is widely believed to be essential for effective therapeutic encounters” and “preserving, justifying, and enhancing trust is the fundamental goal of much of medical ethics, and is a prominent objective in healthcare law and public policy.” (Hall 2005) Notably, it is not only about the patient trusting the doctor:

An additional aspect of trust in medicine is trust that the physician has in the patient. Mutual trust is an important aspect of the patient–physician relationship, with potential consequences for each of the 2 parties. We now know that overt demonstrations of physician’s trust in the patient enhance the patient’s trust of the physician. The opposite is also true; when patients perceive lack of trust by the physician, they tend to diminish their trust in the physician. Mutual trust improves cooperation and reduces the need for monitoring. (Pellegrini 2017, 98)

10. Roberts (2016, 79). Of note, the specific rules and codes of conduct may differ among legal systems and states.



At this point, it may be helpful to distinguish between two types of NTA application. First is acquiring information the person herself could also provide. A type of NTA that falls into this category is lie detection: The person herself could also tell the examiner that she is lying (a person knows when she is lying). The reason for lie detection is that the person's own utterances are not trusted. The second type of NTA concerns acquiring information the person herself could not provide, either because she doesn't know, or because she lacks the physical ability to communicate the information. An example is a person in a functional locked-in<sup>11</sup> situation who cannot express her thoughts unless assisted by NTA. While using NTA in the first context is likely to convey distrust, in the second, it is unlikely to be an issue. Note that in psychiatry, especially in forensic psychiatry, the patient is usually able to communicate about her mental life. In those cases distrust may therefore well be a factor. Given the central role trust plays in medicine—and therefore also in psychiatry—this is a significant issue to consider.

Deception in forensic settings is considered to be a common phenomenon. According to (Young 2014), "the prevalence of malingering in the forensic disability and related context" has been estimated to be a "percentage of up to 50% or so." At present, forensic practitioners can assess deception by evaluating the presence of certain "cues," for instance, rare symptoms, unexpected symptom combinations, and the severity of the symptoms (Rogers and Granacher 2011). In addition, forensic practitioners may use structured tools to assess the possibility of deception and malingering, such as the Structured Inventory of Malingered Symptomatology (SIMS) (Van Impelen et al. 2014). Current skills and tests are far from perfect, however, and deception continues to be an important issue in forensic psychiatric evaluations. NTA could be used to diminish that problem. Yet the use of NTA could be considered an explicit and serious expression of the psychiatrist's distrust of the person.

In sum, the use of NTA in psychiatry can be perceived as potentially undermining trust, which deserves specific attention should it be introduced in this field.

### Consent and Judgmental Ability

Before NTA can be used in health care, in principle, a patient's consent has to be obtained. There are several problems regarding consent for NTA. First, it may be difficult to obtain consent since the implications of

giving it may not be foreseeable.<sup>12</sup> NTA concerns a cutting-edge development and possibilities for future use are enormous, yet hard to foresee. How is one to explain to patients what they are consenting to, if it is difficult to know future implications of such consent? This is a general issue for consent regarding NTA in health care. But in psychiatry, in view of the highly confidential nature of the information psychiatrists are likely to be interested in, and also in view of what has been said about (breaches and limitations of) confidentiality, it is perhaps even more critical that the patient has sufficient information to give valid consent.

Second, in psychiatry, in particular in severe mental illness, decision-making capacity may well be an issue, and therefore it may be problematic to obtain valid informed consent from the person concerned (Owen et al. 2008). Even though incapacity is certainly not limited to severe mental illness, in severe mental illness it is a specific concern (Helmchen 2015). Therefore, this issue is particularly relevant to the psychiatrist using NTA.<sup>13</sup>

After reports that using fMRI made it possible to communicate with patients in a (presumed) vegetative state, the question arose of whether NTA could enable these patients to make treatment decisions, including those regarding end of life.<sup>14</sup> Clearly, this leads to another question: Are these patients able to make such decisions, and could their capacity be assessed using NTA? In clinical practice, psychiatrists are often considered the experts on assessments of competency, and they tend to be consulted in complicated cases (Appelbaum 2007). If decisional capacity would have to be assessed by NTA in a patient who is in a functional locked-in state, this would most probably count as a complicated case. Therefore, psychiatrists may well be involved. For psychiatrists, this creates the intricate question of how to assess judgmental capacity in such a situation. Of note, competency has to be assessed regarding a specific medical decision

11. According to Bruno et al. (2011), "the term *functional locked-in syndrome* could be proposed for patients with a dissociation between extreme behavioural motor dysfunction and the identified preserved higher cognitive functions only measurable by functional imaging techniques." On trust in the NTA context, see also Meynen (2018b).

12. See Roskies (2015), Claydon (2017), and see Yuste et al. (2017) on consent regarding neurotechnologies. This is to some extent similar to approving research on your tissues, where you might not be aware what use might be made of the tissue or research results in the future.

13. See also Kelly (2012) on brain imaging in psychiatry.

14. Demertzi and Laureys (2015, 674) write: "It should be stressed that these exciting developments [concerning NTA in functional locked-in syndrome] are not yet routine in clinical practice. The first obstacle that needs to be overcome before the methodologies discussed above can enter clinical practice relates to ethical concerns. For example, in terms of treatment planning, such as pain management and end-of-life decision making, patients with disorders of consciousness are now offered the possibility to express their preferences by means of brain-computer interfaces. What remains to be clarified is the degree to which such indirect responses can be considered reliable and worthy of legal representation."

(Appelbaum 2007). This implies that a person can be able to make treatment decision A, while unable to make treatment decision B. In addition, a sliding scale regarding decisional capacity is often used (Drane 1985). The more serious the possible negative consequences of a decision, the higher is the threshold for competency. Thus, at least in principle, it might be that a patient who is able to communicate only through NTA is considered to be competent to make a decision about using an ointment for a skin problem, but not to make end of life decisions (in which the stakes are much higher). Both issues, the specificity of the competency assessment and the sliding scale, may further complicate the evaluations.

Another ethical NTA issue concerns coercion: performing NTA in a situation in which the patient withholds consent, or is somehow pressured to give consent.<sup>15</sup> Although not exclusively relevant to mental health care, coercive interventions (such as involuntary commitment and treatment), are particularly relevant for psychiatry. Would coercive use of NTA be acceptable? The first question we have to answer concerns a technical issue: Can the NTA technique be performed in a resisting subject? fMRI is sensitive to the slightest movements,<sup>16</sup> so at present neuroimaging of a reluctant and resisting person in psychiatry appears impossible.

But what about "coerced consent" (Roskies 2015)? One of the central tasks in forensic psychiatry is prediction of recidivism. NTA could become helpful in predicting such risk (Roskies 2015). For instance, a recent study found that it was possible, at least to some extent, to predict future rearrest using fMRI (Aharoni et al. 2013). Currently, prediction of future crimes is still very difficult, and it would be highly valuable to have an additional information source in this respect.<sup>17</sup> Imagine a violent offender who has spent many years in a forensic psychiatric facility, who would like to go on leave or be paroled. Based on classic risk assessment, the psychiatrist believes that the risk is still too high. Nevertheless, NTA might show that the risk of recidivism is actually low. Therefore, the forensic psychiatrist could tell the patient: "Based on our risk assessment, leave is not possible, but NTA *could* show that it is safe. Do you consent to NTA?" After many years in a forensic psychiatric facility, wouldn't that be, in the words of the *Godfather*, "an offer you cannot refuse"? How is one to assess such situations, which are likely to

be more relevant to (forensic) psychiatry than to any other area of medicine?<sup>18</sup>

Meanwhile, we may respond that giving such a choice to a forensic patient is a legitimate offer, not a threat to make that person's position "worse" if he would not consent to NTA. Yet this does not imply that it is irrelevant for the issue of coercion. Szmukler and Appelbaum consider types of coercion in mental health care, and they approach the issue from the perspective of a "spectrum of treatment pressures" (emphasis added). They distinguish between the following forms of pressures (in ascending force): persuasion, interpersonal leverage; inducements; threats; and compulsory treatment. One of the interesting points in their discussion of the range of "pressures" is that it is hard to clearly distinguish between these. To a large extent it depends on one's (moral) conception of the context. Suppose a psychiatrist considers it important that a patient takes the medication. During a conversation with the patient she says: "If you don't take the medication, I will ask for a court order." Is this just transparent information about the procedure, or is it a threat? It could be that the psychiatrist merely intends to inform the patient, but the patient's perception of her words could be different—and that is not irrelevant. Several accounts of coercion have emphasized the subjective part—how the words are experienced by the other person—as relevant to an analysis of coercion (Szmukler and Appelbaum 2008). In a forensic NTA situation as sketched, it seems not unlikely that the patient may perceive the offer as a threat: If you don't consent to NTA, leave is not an option.

The point is that NTA is a new technique, and particularly in psychiatry it could be very "invasive" in nature. Still, people may argue that as long as the patient consents to it, it is acceptable. However, regarding invasive interventions, such as deep brain stimulation (DBS) in a forensic population (Fuss et al. 2015), it has been argued that no hopes (offers, pressures) in terms of leave or parole are admissible. In other words, we have to determine the "invasiveness" of NTA in psychiatry, and determine in what situations and to what extent certain "pressures" are permitted and ethically acceptable.

## Responsibility

Some NTA data may be directly "action guiding." This is the case when a person's decisions (or intentions) can be "apprehended" and translated into action. The translation of those decisions may be mediated by

15. See also Meynen (2017, 2018b) on NTA in the context of forensic psychiatry.

16. See also Roskies (2015) and Ryberg (2017).

17. In this section I focus on the NTA-based prediction of harm against others. But NTA could in principle also help assess the probability of harm against a patient himself, e.g., the risk of suicide (see the already mentioned study by Just et al. 2017).

18. Our ethical evaluation of cases like these could well be related to the invasiveness of the technique. Clearly, if the technique would be physically harmful (e.g., if radioactive material is used), that would be an important issue for ethical consideration (see Roskies 2015; and regarding brain imaging in psychiatry, see Kelly 2012).

other people, such as a decision about treatment through NTA in a patient in a functional locked-in state: The patient's decision may be translated into action by the physician or the nurse. Alternatively, the translation can be performed by a robot through a BCI. This is the case in a paralyzed patient who directs a robotic arm in a certain way. When NTA is in one of these senses "action guiding," it enables a person to control and change the world and, more specifically, his own physical or mental condition. This is likely to be a very powerful way in which NTA can contribute to quality of life in health care.

Ethically, when actions are concerned, the concept of responsibility becomes relevant. When a person performs an action, particularly an action that harms another person, we may ask whether it is justified to hold that person responsible for the action.

But who is responsible for the choice or action the person has initiated through NTA? Recently, the following scenario was sketched in a *Nature* paper:

A paralyzed man participates in a clinical trial of a brain-computer interface (BCI). A computer connected to a chip in his brain is trained to interpret the neural activity resulting from his mental rehearsals of an action. The computer generates commands that move a robotic arm. One day, the man feels frustrated with the experimental team. Later, his robotic hand crushes a cup after taking it from one of the research assistants, and hurts the assistant. Apologizing for what he says must have been a malfunction of the device, he wonders whether his frustration with the team played a part. This scenario is hypothetical. But it illustrates some of the challenges that society might be heading towards. (Yuste et al. 2017)

Of course, there are many possible perspectives on a case like this.<sup>19</sup> Suppose that it turns out that the robotic arm caused some serious damage to the assistant's body. And suppose the matter is considered by a court. The reason for paralysis in this case is unknown, but it might be because of brain damage. The question might come up to which extent the patient's mental capacities are

sufficient to bear responsibility for this action. It is not unthinkable that in such a case, a psychiatric evaluation is requested. The action that harmed the assistant could have been an impulsive action. But was it due to the patient's impulse control problems (e.g., after a stroke), or "impulse control problems" in the BCI-robot-arm system? In other words, was the patient too fast in forming an intention to hit the assistant, or was the BCI too sensitive, picking up a mere thought as being an intention or decision? Where is one to put the blame: in the "brain," or the BCI?

Within the realm of medicine, psychiatrists tend to be consulted where the law is confronted with issues of responsibility related to possibly disturbed mental functioning. In health care, BCI devices are likely to be used in people with serious neurological conditions. Some of these conditions may be due to brain damage. If a scenario as just described were to occur and serious harm were to be a consequence, would the court just assume the patient's mental functioning to be intact, or would, at least in some cases, a psychiatrist be called to make an assessment of the patient and his (brain's) interaction with the device? In such an assessment it may be important to evaluate the extent to which a person could control his or her behavior (in many jurisdictions, this is an important factor in evaluating insanity; Simon and Ahn-Redding 2006). In this case, a relevant question may involve whether the NTA technique detected prereflective mental states (such as inclinations) and whether it translated such a state into action before the patient could reflect on it. To what extent could a person be held responsible for actions resulting from this type of activation of the BCI? In our society, thoughts are considered "free": In principle, people cannot be held responsible for a thought per se. Only when they act on their thoughts can they be held accountable. BCI may not only detect a person's intentions, but it could also translate mental states into action that would not have been translated into action if the person's body had been intact and he did not have to use BCI. But for a forensic psychiatrist, it may be very difficult to evaluate the person's behavioral control in such a case. Here, I merely point at the difficulty of ascribing responsibility in a forensic psychiatric BCI case. It is clear that there may be many more sociolegal and ethical consequences of such techniques regarding responsibility (e.g., concerning the manufacturer), which I leave aside.

### Contextualizing NTA: Ecological Validity

Regarding the use of NTA, either in psychiatry or elsewhere in medicine, it is crucial that the techniques have been sufficiently tested. Using inaccurate methods is never ethical, as is not knowing how to use them and interpret their results. I do not consider here the many technical, methodological, and interpretational pitfalls

19. Mattia and Tamburrini (2015) write on responsibility and liability regarding BCI: "There are ethically relevant issues arising from known epistemic limitations about the behaviors of brain-actuated robots. The retrospective distribution of responsibilities and liabilities for damages caused by BCI-actuated robots is a case in point, insofar as programmers, manufacturers, and users are not in the position to predict exactly and certify what these robots will actually do in their intended operational environments ... For example, how are responsibilities and liabilities sensibly distributed if a BCI-actuated robotic wheelchair rolled down a staircase due to its incorrect interpretation of user intents or when a robotic arm responding to a glass-of-water request hits another person standing in the room?" See Haselager (et al. 2009) on BCI team responsibility, which I do not discuss here.



and caveats that are related to NTA.<sup>20</sup> Yet I have pointed out the relevance of psychiatric contexts of NTA application in medicine. Such emphasis on context entails a specific technological/methodological concern: the ecological validity or the “fit” of a technique to a particular setting.<sup>21</sup>

Let us consider fMRI lie detection. This type of NTA concerns an important area of research, and in the laboratory it appears possible to distinguish lies from truthful statements, at least to some extent (Langleben et al. 2016). Meanwhile, there is the problem of ecological validity: Does the method also work in the context of intended use, for instance, in the criminal justice system?<sup>22</sup> Subjects willing to cooperate with scientific research on fMRI lie detection most probably play by the rules of the experimental design. Defendants, meanwhile, may not be willing to play by the rules and may try to distort or manipulate the results. They may take so-called countermeasures (Wolpe et al. 2005), aimed at distortion and manipulation, and they may actually be successful. Therefore, unless manipulation can be sufficiently detected, or unless the techniques can be made sufficiently immune to such countermeasures, the techniques don’t appear to be ready for courtroom use. In other words, recognizing the ethical priorities of NTA in the specific context of psychiatry should also entail recognizing the importance of the ecological validity of the techniques used. They should not only be valid in the relevant populations (such as people with severe mental illness), but also in the relevant situations, such as stressful circumstances, and those in which the risk of deception is increased (Meynen 2017). Recognizing the limitations of techniques regarding their ecological validity may also help to avoid unwarranted interpretations of the data, such as by an expert in the court of law (Silva 2009). Generally, in my view, the criterion for using NTA technology in clinic and courtroom should be its added value compared to other techniques (Meynen 2016). This implies that there is no absolute standard (such as “eighty-five percent accuracy”), but that its value basically depends on the availability or absence of other (suitable) techniques. Of note, in the courtroom, legal thresholds for admissibility of evidence may well apply—depending on the legal system (in the United States, *Daubert* and *Frye* are relevant in this respect) (Meynen 2016).

20. See also the earlier section on limitations of current NTA techniques and see, e.g., Blitz (2017) and Pardo and Patterson (2013).

21. On the importance of ecological validity of NTA see Roskies (2015).

22. See on this topic Langleben and Moriarty (2013), Rusconi and Mitchener-Nissen (2013), and regarding forensic psychiatry Meynen (2017).

## CONCLUSION

In this article, I explored ethical challenges for possible future psychiatric use of NTA. These challenges regard, first, confidentiality. NTA in psychiatry may concern the most private parts of the mind, which calls for utmost confidentiality safeguards, for example, regarding data storage and access. At the same time, psychiatrists sometimes operate in a context in which confidentiality is limited, especially at the intersection with the law. Furthermore, in case NTA would reveal information about danger, the psychiatrist may have the duty to breach confidentiality. Second, using NTA may be a sign of distrust, while in medicine, and therefore in psychiatry, trust is a core value. NTA can be perceived as an expression of distrust particularly in those cases in which the patient is in principle (physically) able to communicate the information. In psychiatric patients, this ability is usually preserved. Therefore, the issue of distrust and NTA has to be dealt with. Third, obtaining consent may be complicated by the fact that it may be hard for a patient to grasp—and for a doctor to foresee and explain—the possible future implications (and risks) of NTA. In the context of forensic psychiatry, the possibility of coerced consent constitutes a prominent ethical concern. Furthermore, NTA could lead to questions regarding responsibility, in particular regarding BCI and robot-mediated action. Psychiatrists could be consulted to answer such questions. Finally, I addressed the issue of the ecological validity of NTA regarding the context of psychiatry: One should carefully consider whether the technique is methodologically appropriate for the psychiatric context of use. Clearly, this analysis was not meant to be exhaustive and more could be said about each of the above-mentioned issues. It was my aim to show that psychiatric NTA applications entail some specific ethical concerns. Further research would be required to examine them in more detail, and to deal with them.

## ACKNOWLEDGEMENT

I am grateful to Ben Ferguson for his helpful suggestions. ■

## REFERENCES

- Aharoni, E., G. M. Vincent, C. L. Harenski, V. D. Calhoun, W. Sinnott-Armstrong, M. S. Gazzaniga, and K. A. Kiehl. 2013. Neuroprediction of future rearrest. *Proceedings of the National Academy of Sciences of the United States of America* 110(15): 6223–6228. doi: [10.1073/pnas.1219302110](https://doi.org/10.1073/pnas.1219302110).
- Appelbaum, P. S. 2007. Clinical practice. Assessment of patients' competence to consent to treatment. *New England Journal of Medicine* 357(18): 1834–1840. doi: [10.1056/NEJMcp074045](https://doi.org/10.1056/NEJMcp074045).
- Beauchamp, T. L., and J. F. Childress. 2013. *Principles of biomedical ethics*. 7th ed. New York, USA: Oxford University Press.

- Blitz, M. J. 2017. *Searching minds by scanning brains. Neuroscience technology and constitutional privacy protection*. Basingstoke, UK: Palgrave.
- Bruno, M. A., A. Vanhaudenhuyse, A. Thibaut, G. Moonen, and S. Laureys. 2011. From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: Recent advances in our understanding of disorders of consciousness. *Journal of Neurology* 258(7): 1373–1384. doi: [10.1007/s00415-011-6114-x](https://doi.org/10.1007/s00415-011-6114-x).
- Claydon, L. 2017. Brain-based mind reading for lawyers: Reflecting on possibilities and perils. *Journal of Law and the Biosciences* 4(3): 594–598. doi: [10.1093/jlb/lxx032](https://doi.org/10.1093/jlb/lxx032).
- Demertzi, A., and S. Laureys. 2015. Detecting levels of consciousness. In *Handbook of neuroethics*, eds. J. Clausen and N. Levy, 665–677. Dordrecht, Netherlands: Springer.
- Drane, J. F. 1985. The many faces of competency. *The Hastings Center Report* 15(2): 17–21.
- Evers, K. and Sigman, M. 2013. Possibilities and limits of mind-reading: A neurophilosophical perspective. *Consciousness and Cognition* 22, 887–897.
- Feuerstein, S., V. Coric, F. Fortunati, S. Southwick, H. Temporini, and C. A. Morgan. 2005. Malingering and forensic psychiatry. *Psychiatry (Edmont)*, 2(12): 25–28.
- Fuchs, T. 2006. Ethical issues in neuroscience. *Current Opinion in Psychiatry* 19(6): 600–607. doi: [10.1097/01.yco.0000245752.75879.26](https://doi.org/10.1097/01.yco.0000245752.75879.26).
- Fuss, J., M. K. Auer, S. V. Biedermann, P. Briken, and W. Hacke. 2015. Deep brain stimulation to reduce sexual drive. *Journal of Psychiatry & Neuroscience* 40(3): 150003.
- Gutheil, T. G. 2005. Ethics and forensic psychiatry. In *Psychiatric ethics, third edition*, eds. S. Bloch, P. Chodoff, and S. A. Green, 345–361. New York, USA: Oxford University Press.
- Hall, M. A. 2005. The importance of trust for ethics, law, and public policy. *Cambridge Quarterly of Healthcare Ethics* 14(2): 156–167.
- Haselager, P., R. Vlek, J. Hill, and F. Nijboer. 2009. A note on ethical aspects of BCI. *Neural Networks* 22(9): 1352–1357. doi: [10.1016/j.neunet.2009.06.046](https://doi.org/10.1016/j.neunet.2009.06.046).
- Helmchen, H. 2015. Ethics in psychiatry. In *Handbook of neuroethics*, ed. J. Clausen and N. Levy, 873–878. Dordrecht, Netherlands: Springer.
- Ienca, M., and R. Andorno. 2017. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy* 13(1): 5. doi: [10.1186/s40504-017-0050-1](https://doi.org/10.1186/s40504-017-0050-1).
- Just, M. A., L. Pan, V. L. Cherkassky, D. McMakin, C. Cha, M. K. Nock, and D. Brent. 2017. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature Human Behaviour* 1(12): 911–919. doi: [10.1038/s41562-017-0234-y](https://doi.org/10.1038/s41562-017-0234-y).
- Kelly, B. D. 2012. Brain imaging in clinical psychiatry: Why? In *I know what you're thinking: Brain imaging and mental privacy*, ed. S. Richmond, G. Rees, and S. Edwards, eds. 111–121. Oxford, UK: Oxford University Press.
- Langleben, D. D., J. G. Hakun, D. Seelig, A. L. Wang, K. Ruparel, W. B. Bilker, and R. C. Gur. 2016. Polygraphy and functional magnetic resonance imaging in lie detection: A controlled blind comparison using. *The Journal of Clinical Psychiatry* 77(10): 1372–1380. doi: [10.4088/JCP.15m09785](https://doi.org/10.4088/JCP.15m09785).
- Langleben, D. D., and J. C. Moriarty. 2013. Using brain imaging for lie detection: Where science, law and research policy collide. *Psychology, Public Policy and Law* 19(2): 222–234. doi: [10.1037/a0028841](https://doi.org/10.1037/a0028841).
- Leroy, A., J. R. Foucher, D. Pins, et al. 2017. fMRI capture of auditory hallucinations: Validation of the two-steps method. *Human Brain Mapping* 38(10): 4966–4979. doi: [10.1002/hbm.23707](https://doi.org/10.1002/hbm.23707).
- Linden, D. 2012. Overcoming self-report: Possibilities and limitations of brain imaging in psychiatry. In *I know what you're thinking: Brain imaging and mental privacy*, eds. S. Richmond, G. Rees, and S. Edwards, 123–135. Oxford, UK: Oxford University Press.
- Mason, R. A., and M. A. Just. 2016. Neural representations of physics concepts. *Psychological Science* 27(6): 904–913. doi: [10.1177/0956797616641941](https://doi.org/10.1177/0956797616641941).
- Mattia, D., and G. Tamburrini. 2015. Ethical issues in brain-computer interface research and systems for motor control. In *Handbook of neuroethics*, eds. J. Clausen and N. Levy. Dordrecht, the Netherlands: Springer.
- Meynen, G. 2016. *Legal insanity. Explorations in psychiatry, law, and ethics*. Berlin, Germany: Springer.
- Meynen, G. 2017. Brain-based mind reading in forensic psychiatry: Exploring possibilities and perils. *Journal of Law and the Biosciences* 4(2): 311–329. doi: [10.1093/jlb/lxx006](https://doi.org/10.1093/jlb/lxx006).
- Meynen, G. 2018a. Author's Response to Peer Commentaries: Brain-based mind reading: Conceptual clarifications and legal applications. *Journal of Law and the Biosciences* 5(1): 212–216.
- Meynen, G. 2018b. Forensic psychiatry and neurolaw: Description, developments, and debates. *International Journal of Law and Psychiatry*. doi: [10.1016/j.ijlp.2018.04.005](https://doi.org/10.1016/j.ijlp.2018.04.005).
- Monti, M. M., A. Vanhaudenhuyse, M. R. Coleman, M. Boly, J. D. Pickard, L. Tshibanda, A. M. Owen, and S. Laureys. 2010. Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine* 362(7): 579–589. doi: [10.1056/NEJMoa0905370](https://doi.org/10.1056/NEJMoa0905370).
- Nadelhoffer, T., and W. Sinnott-Armstrong. 2012. Neurolaw and neuroprediction: Potential promises and perils. *Philosophy Compass* 7(9): 631–642. doi: [10.1111/j.1747-9991.2012.00494.x](https://doi.org/10.1111/j.1747-9991.2012.00494.x).
- Owen, G. S., G. Richardson, A. S. David, G. Szukler, P. Hayward, and M. Hotopf. 2008. Mental capacity to make decisions on treatment in people admitted to psychiatric hospitals: Cross sectional study. *BMJ* 337(jun30 1): a448. doi: [10.1136/bmj.39580.546597.BE](https://doi.org/10.1136/bmj.39580.546597.BE).
- Pardo, M. S., and D. Patterson. 2013. *Minds, brains, and law. The conceptual foundations of law and neuroscience*. New York, USA: Oxford University Press.

- Pellegrini, C. A. 2017. Trust: The keystone of the patient-physician relationship. *Journal of the American College of Surgeons* 224(2): 95–102. doi: [10.1016/j.jamcollsurg.2016.10.032](https://doi.org/10.1016/j.jamcollsurg.2016.10.032).
- Richmond, S., G. Rees, and S. J. L. Edwards. (Eds.) 2012. *I know what you're thinking Brain imaging and mental privacy*. Oxford, UK: Oxford University Press.
- Roberts, L. W. 2016. *A clinical guide to psychiatric ethics*. Arlington, VA: American Psychiatric Association.
- Rogers, R. (Ed.) 2012. *Clinical assessment of malingering and deception*. New York, USA: The Guilford Press.
- Rogers, R., and R. P. Granacher. 2011. Conceptualization and assessment of malingering. In *Handbook of forensic assessment: Psychological and psychiatric perspectives*, eds. E. Y. Drogin, F. M. Dattilio, R. L. Sadoff, and T. G. Gutheil. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Roskies, A. L. 2015. Mind reading, lie detection, and privacy. In *Handbook of neuroethics*, eds. J. Clausen and N. Levy, 679–695. Dordrecht, Netherlands: Springer.
- Rusconi, E., and T. Mitchener-Nissen. 2013. Prospects of functional magnetic resonance imaging as lie detector. *Frontiers in Human Neuroscience* 7: 594. doi: [10.3389/fnhum.2013.00594](https://doi.org/10.3389/fnhum.2013.00594).
- Ryberg, J. 2017. Neuroscience, mind reading and mental privacy. *Res Publica* 23(2): 197–211. doi: [10.1007/s11158-016-9343-0](https://doi.org/10.1007/s11158-016-9343-0).
- Shen, F. X. 2013. Neuroscience, mental privacy, and the law. *Harvard Journal of Law & Public Policy* 36: 653–713.
- Silva, J. A. 2009. Forensic psychiatry, neuroscience, and the law. *The Journal of the American Academy of Psychiatry and the Law* 37(4): 489–502.
- Simon, R. J., and H. Ahn-Redding. 2006. *The insanity defense, the world over*. Lanham, MD: Lexington Books.
- Soekadar, S. R., N. Birbaumer, M. W. Slutzky, and L. G. Cohen. 2015. Brain-machine interfaces in neurorehabilitation of stroke. *Neurobiology of Disease* 83: 172–179. doi: [10.1016/j.nbd.2014.11.025](https://doi.org/10.1016/j.nbd.2014.11.025).
- Szmukler, G., and P. S. Appelbaum. 2008. Treatment pressures, leverage, coercion, and compulsion in mental health care. *Journal of Mental Health* 17(3): 233–244. doi: [10.1080/09638230802052203](https://doi.org/10.1080/09638230802052203).
- Van Impelen, A., H. Merckelbach, M. Jelicic, and T. Merten. 2014. The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist* 28(8): 1336–1365. doi: [10.1080/13854046.2014.984763](https://doi.org/10.1080/13854046.2014.984763).
- Wolpe, P. R., K. R. Foster, and D. D. Langleben. 2005. Emerging neurotechnologies for lie-detection: Promises and perils. *The American Journal of Bioethics* 5(2): 39–49. doi: [10.1080/15265160590923367](https://doi.org/10.1080/15265160590923367).
- Young, G. 2014. *Malingering, feigning, and response bias in psychiatric/psychological injury*. Dordrecht, Netherlands: Springer.
- Yuste, R., S. Goering, B. A. y Arcas, et al. 2017. Four ethical priorities for neurotechnologies and AI. *Nature* 551(7679): 159–163. doi: [10.1038/551159a](https://doi.org/10.1038/551159a).