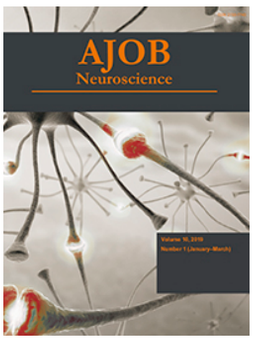

AJOB Neuroscience

ISSN: 2150-7740 (Print) 2150-7759 (Online) Journal homepage: https://www.tandfonline.com/loi/uabn20

The Ethics of Counting Neural Activity as Proof

Annemarie van Stee & Marc Slors

To cite this article: Annemarie van Stee & Marc Slors (2019) The Ethics of Counting Neural Activity as Proof, AJOB Neuroscience, 10:1, 15-16, DOI: 10.1080/21507740.2019.1595782

To link to this article: https://doi.org/10.1080/21507740.2019.1595782

Published online: 09 May 2019.

AJOB Neuroscience, 10(1): 15–34, 2019

ISSN: 2150-7740 print / 2150-7759 online
DOI: 10.1080/21507740.2019.1595782

Open Peer Commentaries

The Ethics of Counting Neural Activity as Proof

Annemarie van Stee, Radboud University
Marc Slors, Radboud University

In "Ethical Issues to Consider Before Introducing Neurotechnological Thought Apprehension in Psychiatry," Gerben Meynen (2019) argues "that normative reflection should be in advance of [neurotechnological] developments" (5). We agree. Meynen notes that so-called neurotechnological thought apprehension (NTA) currently has limitations, as multiple-choice designs are used and outcomes are probabilistic. He notes too that the term "thoughts" in fact covers a wide range of mental states. On these points we agree too. Yet Meynen's list of ethical issues does not take these limitations into account. His normative reflections pertain to the ideal thought experiment that is NTA, not to a realistic perspective on what information we may derive from neuroscientific data, given the technological options we currently have and the directions in which these are developing. Meynen thereby misconstrues certain ethical issues and overlooks one that is both urgent and fundamental: How should we weigh NTA outcomes against other sources of insight into someone's psyche?

Meynen discusses a much too diverse set of studies under the heading "thought apprehension," not all of which are equally likely to become relevant for (forensic) psychiatry. Experiments that are suggestive of "apprehending thoughts," such as communicating with patients in permanent vegetative state, involve instructions that limit the infinite possibilities of real-life thought to very few tractable ones. Brain–computer interfaces (BCIs) are perhaps even more suggestive, but they are not, in fact, mind-reading devices. After extensive training, BCIs pick up brain signals for the control of communication devices and robot arms. However, it is not so much the BCI that is trained to read the patient's brain correctly, but it is the patient who learns to control the BCI. The studies that are most relevant for (forensic) psychiatry aim to assist in predicting future behavior, providing information that people would not be able or willing to disclose otherwise. The studies on predicting suicidality and predicting future rearrest are cases in point. Such predictions, however, are and will remain probabilistic. More importantly, probabilities are inversely related to ecological validity of the predictions.

For example, Marcel Just and colleagues report a 91% accuracy rate with which a classifier can distinguish suicidal youth from nonsuicidal youth on the basis of their functional magnetic resonance imaging (fMRI) patterns (Just et al. 2017). Participants read death-related and life-related words on a screen. They are instructed to think of the properties of that concept for 3 seconds at a time, and think of the same properties each of the six times the word appears. Needless to say, this is very different from thinking "in the wild." These highly artificial conditions require full cooperation and concentration from participants.

What is more, participants in this study belong to two clearly distinguished groups: participants who admit to having current suicidal ideation, and participants who are healthy and have no personal or even family history of psychiatric disorder or suicide attempt. None belong to the group about whom predictions would actually be informative: people who may or may not be suicidal. Just and colleagues selected 34 (out of an original 79) participants with the best data quality, and trained a classifier on data of 33 out of 34 participants. That classifier is able to predict the 34th participant with 91% accuracy. Training a classifier on 18 out of these 34 participants and predicting the group membership of the remaining 16 participants leads to an accuracy rate of 76%. Likewise, a more realistic and therefore varied sample of participants would also lead to a lower accuracy rate.

Now take a study that has higher ecological validity, that is, the study on predicting future rearrest by Eyal Aharoni et al. (2013). Participants were incarcerated men who were about to be released. They had to perform a Go versus No Go task (GNG task), a commonly used task indicating participants' ability to inhibit themselves when the situation asks for it. Aharoni and colleagues followed participants upon their release. They found a statistically significant correlation on the group level

Address correspondence to Annemarie van Stee, Department of Philosophy, Radboud University, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands. E-mail: a.vanstee@ftr.ru.nl

between activity in the anterior cingulate cortex (ACC) during the GNG task and time to rearrest. Approximately 53% of participants were rearrested in the years immediately after their release. Of those who displayed high ACC activity, 46% were rearrested. Of those who displayed low ACC activity, 60% were rearrested. That may be a statistically significant difference, but, as Aharoni and colleagues themselves also stress, it is a far cry from predicting whether a certain individual will be rearrested or not.

MRI technology may become better and classifiers are likely to become more sophisticated. Yet individual-level predictions will always be probabilistic. They will always require the cooperation of participants, who just have to move their heads ever so slightly, or not follow the demanding experimental protocol exactly, for their data to be severely compromised. What is more, high probabilities will always be inversely related to ecological validity of predictions. Meynen's analysis of ethical issues should take these facts as a starting point. Yet at best, he notes them as afterthoughts. With this he risks misconstruing ethical issues.

For example, Meynen sketches a future situation in which a "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?'" (10). This situation will never arise, however, as a classifier will never be able to determine with certainty that someone will not engage in criminal behavior again. Instead, it will classify someone as belonging to a low-risk group about which it has been right in the past in, say, 73% of cases. A premise that underlies Meynen's sketch of the future is that when NTA contradicts risk assessment based on behavior and report, we would rely on NTA. Yet this is exactly the million-dollar question: Given that NTA always leads to probabilistic outcomes, when do we deem it reliable *enough*? Currently, NTA researchers can only make individual-level predictions about recidivism that are not much better than flipping a coin. If sensitivity and specificity levels of predictions were to go up in the future, we might use NTA to bolster our judgments when its outcomes align with other sources of insight into recidivism risk. However, for us to rely on NTA-based predictions when these contradict other sources of insight may always remain a thought experiment (see also Aharoni et al. 2013, 2).

So when do we count brain-based predictions as proof? Again, Meynen has a note on this, an afterthought entirely at the end of the body of his article:

> Generally, in my view, the criterion for using NTA technology in clinic and courtroom should be its added value compared to other techniques … 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. (12)

"Added value" may appear to be a technical issue, a question of when NTA outperforms psychosocial tests with respect to sensitivity and specificity, or whether predictions can be improved by adding NTA as a source of information. Yet, as Meynen rightfully mentions several times, (forensic) psychiatrists may be most interested in NTA insofar as it provides access to the minds of people who are uncooperative and whose reports they do not trust. There is a clear risk that we lower standards of proof for those we deem unreliable. Using "better safe than sorry" arguments, or a version of Meynen's argument that this is all we have given the absence of other reliable evidence, we may deem an accuracy rate of, say, 85% good enough as a basis for judging psychiatric patients or criminal suspects, whereas we deem the matching error rate of 3 out of 20 too high for ordinary people. Violations of "innocent unless proven guilty" may thus hide themselves behind a seemingly technical number.

What is more, relying on brain-based predictions rather than report or clinical impressions is itself an intervention with consequences. The breach of trust that Meynen rightfully notes it involves and its negative effects on the therapeutic relationship between psychiatrist and patient may contribute to a lack of improvement on the part of the patient. Given the poor ecological validity of reliable predictions and the low reliability of predictions that are more ecologically valid, behavioral studies and psychiatrists' impressions are likely to provide better insight for a long time to come, without the breach in trust involved in NTA.

Normative reflection on future developments is notoriously hard. We believe it should start from a thorough acquaintance with the current state of technology and the direction in which technology is developing. The dream of simply "register[ing] all thoughts that come up in a particular period of time" (Meynen 2019, 8), for example, is nothing more than just that: a dream. Meynen has notes and afterthoughts, but does not integrate these into his thinking about the ethical issues. For those, he relies on an ideal thought experiment that will not come to pass. ■

REFERENCES

Aharoni, E., G. M. Vincent, C. L. Harenski, V. D. Calhoun, W. Sinnott-Armstrong, M. S. Gazzaniga, and A. K. En Kent. 2013. Neuroprediction of future rearrest. *Proceedings of the National Academy of Sciences* 110(15): 6223–8. doi: 10.1073/pnas.1219302110.

Just, M. A., L. Pan, V. L. Cherkassky, D. L. McMakin, C. Cha, M. K. Nock, and B. En David. 2017. Machine Learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature Human Behaviour* 1(12): 911–19. doi: 10.1038/s41562-017-0234-y.

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