TRAMES, 2011, **15**(**65/60**), 2, 204–212

BRAIN READING AND MENTAL PRIVACY

Valtteri Arstila[1] and Franklin Scott[2]

[1]University of Turku and New York University, and [2]University of Cincinnati

Abstract. The widespread use and novel applications of brain imaging techniques seem to open the possibility of new threats to one's privacy. Being in a situation where we cannot control what information about ourselves is available to others could restrict the jobs we can get, the business we conduct, the way we are seen by strangers, and the way we relate to our friends and family. While researchers and practitioners should be cautious about the ways they use brain imaging data, we argue that brain reading does not violate privacy in any way different from the already established psychological methods to determine mental phenomena, such as whether someone suffers from color blindness or is clinically depressed. For brain reading to constitute a new threat, we claim it would have to be possible to easily gather information about a subject's mental states in accidental or malicious ways without the knowledge of the subject. Against this possibility, we note that brain imaging techniques require (i) the researchers to intentionally seek a specific type of mental state to the exclusion of information about other states, (ii) the active cooperation and participation of subjects, and (iii) a method of analysis that depends on the already established psychological tests. Hence, while there should be policies that specify when brain imaging can be used and how, these policies should be treated within a broader context of privacy issues in psychology rather than as a special case.

Keywords: brain imaging, fMRI, neuroethics, mental privacy, neuroscience, research ethics, bioethics

DOI: 10.3176/tr.2011.2.08

1. Introduction

Imagine you are applying for a job at an investment bank. You have the education, a hard-working demeanor, and relevant experience from your previous positions. After several rounds of interviews, you make it onto the shortlist and the company wants you to undergo extensive physical and psychological testing. Being relatively fit, you are healthy enough to pass the physical tests. Psychological tests likewise show that you are apt for the open position. For example,

your brain imaging data show that you are unlikely to make financially rash decisions (Kuhnen and Knutson 2005) – which is good when you are responsible for handling other people's money. Yet, you do not get the job because the brain imaging results also revealed some information about you that you would not have wanted to share with the company. This could be, for example, information about your political preferences or how religious a person you are – things unrelated to your work performance that nevertheless make the person hiring new employees choose someone else.

Some argue that this gloomy picture could become reality (e.g. Eaton and Illes 2007, Farah et al. 2008, and Sententia 2004). Technologically it certainly could be: beyond just providing insights on how normal and abnormal neural mechanisms function and on neurological diseases, the majority of research deploying the use of brain imaging techniques is nowadays not confined to medicine. Examples include studies related to our political preferences (Amodio et al. 2007), how social hierarchy manifests itself in the brain activations (Zink et al. 2008), unconscious racial attitudes (Phelps et al. 2000), and lie detection (e.g. Kozel et al. 2005; Sip, Roepstorff and Frith 2008). Indeed, some studies show that brain activation differs between people who have just fallen in love, have been in love for some time, and those who are in love but have been recently been romantically rejected (Aron et al. 2005 and Fisher et al. 2006). None of these studies have much therapeutic or diagnostic value.

What these examples have in common is that they extend the use of brain imaging techniques to complex interpersonal phenomena that are present in our everyday life. The fact that neuroscience is no longer limited to mere medical research is also demonstrated by the prominence it receives in the increasing number of articles in the popular press, many of which un-cautiously suggest the idea of scientists being able to read our minds by the means of brain imaging techniques (see for example Nelson 2010).

As the examples above illustrated, these new possible uses for brain imaging raise the possibility of abusing one of our most fundamental rights: the right to privacy. The worry here is of course the threat of losing one's autonomy – being in a situation where we cannot control what information about ourselves is available to others can restrict the jobs we can get, the business we conduct, the way we are seen by strangers, and the way we relate to our friends and family. Such a situation might occur, for instance, when testing an applicant for certain jobs as in the example above, when enrolling in a health insurance program where we are willing to disclose some information but not everything, or even when being questioned by the police.

The worry about loss of privacy is intensified if this information does not relate to the work or insurance policy we are applying for or to the relevant police investigation (Fuchs 2006). That is, the revealed information might be something that does not have any legitimate function for the employers, insurance companies, and so forth but nevertheless could harm one's life. Our sexual preferences, for example, should not matter to our employers because it does not affect one's

ability to adequately do one's job. Thus if we choose not to disclose it, they should not know it; especially because we might, as a result, be subject to a range of overt or covert discriminations. Further worry is that in many cases a subject might not even know what information is revealed. For example, when someone flashes images of Caucasian and African-American people to you, how would you know whether the person administering the test is determining whether you have some unconscious racial biases based on your amygdala activation rather than a test for normal visual functioning? (Phelps et al. 2000)

Considerations such as those above have been used to raise awareness of the ethical challenges related to brain imaging and threats that they pose for privacy (e.g. Farah et al. 2008, Fuchs 2006, and Sententia 2004). Attributing these challenges and threats to brain imaging are mis-targeted however. Or so we argue below based on the observation that brain reading does not violate privacy in any way different from already established psychological methods to determine mental phenomena, such as whether someone suffers from color blindness or is clinically depressed. Hence, while there should be policies that specify when brain imaging can be used and how, these policies should be treated within a broader context of privacy issues in psychology rather than as a special case.

2. When would brain reading pose a new threat to privacy?

One crucial thing about the brain reading of a mental phenomenon is that, in the final analysis, its predictive power depends on already established psychological tests. For example, Elizabeth Phelps (2000) found that subjects with unconscious racial biases show higher levels of amygdala activation when presented with pictures of unfamiliar African-American faces. Making the claim that there is a correlation between amygdala activation and unconscious racial biases depends on the antecedent soundness of the psychological tests for unconscious racist biases. Likewise, the Kuhnen and Knutson (2005) study showing the correlation between the activation in nucleus accumbens and the tendency to make risk-seeking mistakes depends on the psychological tests for the risk aversiveness being sound. Given such dependency on the psychological tests, brain reading by and large does not pose new threat to the privacy of our mental lives by making available phenomena that could not be revealed without it. That is, we can already violate privacy by the means of traditional psychological tests and brain imaging merely provides new ways to do it.[1]

Given that brain reading is not used to investigate something that could not be investigated by some other means too, the new threat must come from the possible use of these methods that differ from those of the traditional psychological tests.

[1] At the same time it also could be justifiable for a company to run some tests for their prospective employees (as in the example case above, where it might be beneficial to determine a tendency to take risks).

Here, the following three cases, also briefly mentioned above, appear particularly distressing. First, brain imaging could enable researchers to investigate something without the subject knowing what things researchers look for. While this can certainly happen, it is important to notice that the same applies to traditional psychological tests too. At least some features of our personality can be determined without brain imaging and without us even knowing that anything is happening. For instance, it has been found that in certain visual tasks homosexual men perform in the same way as heterosexual women (and unlike heterosexual men) even when the used stimuli is masked in a way that the subjects are not aware that any stimulus was shown (Jiang et al. 2006). Hence, the fact that the subject does not always know what the researchers are looking for does not distinguish brain imaging and psychological tests.

Second, it could be argued that while we can control the information we convey in the traditional psychological tests, we cannot do so in brain imaging tests – that is, only in brain imaging experiments could something that a subject wants to keep as a secret be revealed to researchers. Yet, there hardly is any truth to this. On the one hand, in accordance with the previous point, how can you control what information you reveal about yourself if you do not even know what the researchers are looking for? On the other hand, in many cases we need to be aware of something and be compliant subjects in order for the brain imaging methods to work. For example, as will be discussed in the following section, if a person does not want to disclose certain thoughts or memories and manages not to think about them, then even if we assume that such states could be in principle determined, in practice they cannot be read from the brain imaging data.

Finally, as our example suggested, there could be cases of brain imaging where some information is revealed 'accidentally'. If this is indeed the case, then brain reading would indeed bring about a new threat to privacy because something about the subjects could be revealed without anybody really wanting these secrets to be revealed. This would, in turn, lead to the situation where shame would become a more common feeling in our society (Räikkä 2010), people could come to distrust physicians and scientists causing them to avoid important medical care, or people could feel compelled to forgo certain activities for fear of being discovered (e.g. reading unpopular political literature or engaging in a disapproved of romantic relationship). In the next section, however, we will argue that brain imaging does not reveal our secrets accidentally – that there are no unintended violations of privacy.

Before elaborating on this issue further, it is useful to note the distinction between two notions of transparency of mental phenomena for brain reading. According to the weaker notion, we can (in the future) determine all mental phenomena based on the suitable brain imaging data. According to the stronger notion, we can (in the future) determine all mental phenomena based on the brain imaging data as easily as we can determine different features in the pictures – one research paradigm and one method of analysis would reveal all the secrets. Both cases pose a threat to the privacy of our mental life. They differ; however, in how

significant the threat is because only if the stronger notion is empirically sound, do brain imaging techniques pose a kind of ethical threat to the privacy that differs from the already existing threats. This stronger notion of the transparency is the one we question in the next section, leaving aside the weaker one.

3. Violations of privacy are intentional

The idea that brain imaging data could reveal something unintended – something that the researchers did not look for in the first place – has been recently put forward by Margaret Eaton and Judy Illes (2007). While this obviously does hold for structural characteristics of our cortex where abnormalities are often found in MRI scanning (Illes et al. 2004), the generality of such claim is questionable. It is particularly doubtful in the cases where researchers aim at investigating more 'dynamic' states, such as thoughts, memories, and personality traits, because even accepting weak transparency, discovering one of these states requires research tasks and methods of analysis that are unlikely to reveal some other detailed personal piece of information.

To begin with, most of the current methods rely on contrast between two or more conditions. For example, one typical kind of task asks a subject to recall a particular memory in some trials and a different memory (or no particular memory) in others. The subsequent analysis is then based on contrasting the brain imaging data obtained from both conditions and filtering away the 'noise' and background brain activity. While the task given by the neuroscientist can establish a proper contrast for that phenomenon, it does not extend to many others. Thus the brain imaging data does not contain information that could be used to reveal unintended mental phenomena. For instance, you cannot establish whether a person has unconscious racist biases if you do not show this person stimuli of people from different races.

The requirement that one must establish suitable contrastive conditions also means that some phenomenon cannot be investigated without the cooperation or knowledge of the subject. For example, one cannot determine unconscious thoughts and memories of a subject by using brain imaging methods because if they remain always unconscious, then there is no contrast for the subsequent analysis to later reveal. Likewise, if the subject is reluctant to follow the instructions by, for example, failing to recall the proper memory or by thinking about it when she is not supposed to think about it, then the required contrast also cannot be established.

Second, even if raw brain imaging data do contain information about various extraneous mental phenomena, it is only when researchers bring to bear a method of analysis to that data could light be shed on those phenomena. After all, the brain imaging data consists simply of a great amount of numbers and researchers have to ask 'the right questions', to use the correct analytical methods, to uncover the phenomenon that is behind those numbers. Accordingly, it is doubtful that when

one investigates, say, the political preferences of a person, those same methods would show a person's recent memories of being on vacation.

One issue concerning the analysis is the temporal resolution under investigation. The first stages of visual processing, for example, process information at a much faster rate than later stages (Holcombe 2009), which in turn process information faster than, say, our ability to shift attention from one place to another (Egeth and Yantis 1997). Hence, when one investigates how well we can recognize words flashed to the screen (a task the US military uses to test newly developed pilot's helmets), the method of analyzing the brain imaging data does not reveal anything about the earlier processing stages (because the time course used for the segmenting the data is too coarse) nor does this method reveal anything about the allocation of attention (because the time course used for the segmenting is too fine). This does not mean that the same brain imaging data could not be used to investigate those phenomena too – merely that investigating those requires a different, separate method of analysis. Hence information concerning them is not revealed by accident.

A possibly more important issue concerns the brain area under investigation. Brains are, in a sense, modular: although different brain areas are heavily interconnected, each of these areas also processes some specific kind of information that other areas do not process. To give a few examples, damage to the ventromedial prefrontal cortex produces utilitarian biases in making moral judgments (Koenigs et al. 2007), whereas the amygdala activation relates to the processing of certain emotions (Zald 2003), and numerical information is processed (in addition to working memory) in the mid-frontal area and intraparietal sulcus (Nieder and Miller 2004).

What this means for the analysis of brain imaging data is that in many cases the method of analyzing simply ignores data from the parts of the brain that do not process the kind of information that is under investigation. Significantly, ignoring this information may even increase the accuracy of the analysis. Eleanor Maguire's research team studying working memory found that after several steps in machine learning, the researchers were able to predict with greater than chance probability which of the three films the subject was reminiscing about using only imaging data (Chadwick et al. 2010). Importantly for the issue at hand, the accuracy of their method increased once they focused on the brain imaging data from the hippocampus, which is known to process memories. Because this method only focuses on brain activation in the hippocampus, it could not reveal any unintended phenomena that do not depend on the activation in this area.

Given the specificity of the research tasks and the method of analysis of brain imaging, the violations of privacy are in practice always intentional. So, brain reading is not as easy as simply glancing at a photograph; brain imagining studies do not make our mental states strongly transparent. Rather, the threat to mental privacy from brain reading is no different from the threat posed by other forms of psychological testing.

4. Further considerations

The purpose of this paper was to shed light on the way our autonomy and mental privacy is threatened by the recent advancements in brain imaging techniques. Our conclusion was that these advancements do not bring about new ethical challenges for privacy because brain imaging is used to investigate phenomena that can be investigated by other means too and because the violations of privacy occur in the same way in traditional psychological tests as with brain imaging experiments. This does not mean that violating privacy in either of the methods is justifiable. It means, however, that brain imaging techniques do not bring about any new kind of threats for privacy.[2]

It is worth distinguishing the discussion to this point about using brain imaging to violate one's privacy from a related issue. We have described various ways brain imaging data are collected and analyzed in what's today considered to be scientifically sound ways. A further problem is the possibility of people analyzing brain imaging data in unsound ways, making inferences that are not justified by the data and not recognizing the possibility of error or exceptions to generalizations. The brain is plastic and even if activation in a particular brain area indicates some particular mental phenomenon in a large percentage of the population, we should expect there to be exceptions. Just as the popular press makes brain imaging studies seem far more breathtaking, certain, and unqualified than they really are, researchers examining particular individuals could fall into the same traps (especially if they have some agenda or stake in the outcome). Because brain imaging analysis is complicated and relatively new, such cases are perhaps more likely than similar problems with psychological tests. While serious, we do not think these kinds of cases ought to be considered a violation of mental privacy, though, because the unsound claims do not reveal anything true about the subject's mental states. It would be some other kind of violation of autonomy, more akin to defamation.

Overall, in light of the increasingly widespread use and application of brain imaging studies, it would be wise to have discussions concerning the acceptable uses of different tests that eventually lead to commonly accepted policies – for example whether employers should conduct psychological tests for their prospective applicants without much regulation.[3] At the same time, it is not obvious that this issue is very pressing: given that different psychological tests have been around for some time and have been amenable to reasonable polices controlling

---

[2] This does not mean that brain imaging would not bring about new ethical challenges that are not related to privacy though. Quite the contrary, such new ethical challenges are likely to occur.

[3] There is, of course, the code of ethics adopted by the American Psychological Association – and the similar associations in other countries – as well as laws that prevent discriminating applicants based on their sexual preferences, religious views, and so forth. Nevertheless, these guidelines and laws do not prevent testing such issues per se if one wants to, say, for the purpose of integrity testing.

their use. Furthermore, it is not obvious why such things would change if similar or novel mental phenomena are revealed by the use of brain imaging methods.

Addresses:
Valtteri Arstila
Department of Behavioral sciences and philosophy
University of Turku
20014 Turku
Finland
E-mail: valtteri.arstila@utu.fi

Franklin Scott
Department of Philosophy
University of Cincinnati
PO Box 210374
Cincinnati, OH 45221-0374, USA
E-mail: franklin.scott@uc.edu

References

Amodio, D. M., J. T. Jost, S. L. Master, and C. M. Yee (2007) "Neurocognitive correlates of liberalism and conservatism". *Nature Neuroscience* 10, 10, 1246–1247.

Aron, A., H. Fisher, D. J. Mashek, G. Strong, H. Li, and L. L. Brown (2005) "Reward, motivation, and emotion systems associated with early-stage intense romantic love". *Journal of Neurophysiology* 94, 1, 327–337.

Chadwick, M. J., D. Hassabis, N. Weiskopf, and E.A. Maguire (2010) "Decoding individual episodic memory traces in the human hippocampus". *Current Biology* 20, 6, 544–547.

Eaton, M. L., and J. Illes (2007) "Commercializing cognitive neurotechnology – the ethical terrain". *Nature biotechnology* 25, 393–397.

Egeth, H. E., and S. Yantis (1997) "Visual attention: control, representation, and time course". *Annual Review of Psychology* 48, 269–297.

Farah, M. J., M. E. Smith, C. Gawuga, D. Lindsell, and D. Foster (2008) "Brain imaging and brain privacy: a realistic concern?". *Journal of Cognitive Neuroscience* 21, 119–127.

Fisher, H. E., A. Aron, and Lucy L. Brown (2006) "Romantic love: a mammalian brain system for mate choice". *Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences* 361, 1476, 2173–2186.

Fuchs, T. (2006) "Ethical issues in neuroscience". *Current Opinion in Psychiatry* 19, 600–607.

Holcombe, A. O. (2009) "Seeing slow and seeing fast: two limits on perception". *Trends in Cognitive Sciences* 13, 5, 216–221.

Illes, J., A. C. Rosen, L. Huang, R. A. Goldstein, T. A. Raffin, and G. Swan (2004) "Ethical consideration of incidental findings on adult brain MRI in research". *Neurology* 62, 888–890.

Jiang, Y., P. Costello, F. Fang, M. Huang, and S. He (2006) "A gender- and sexual orientation-dependent spatial attentional effect of invisible images". *Proceedings of the National Academy of Sciences of the United States of America* 103, 45, 17048–17052.

Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio (2007) "Damage to the prefrontal cortex increases utilitarian moral judgments". *Nature* 446, 7138, 908–911.

Kozel, F. A., K. A. Johnson, Q. Mu, E. L. Grenesko, S. J. Laken, and M. S. George (2005) "Detecting deception using functional magnetic resonance imaging". *Biological Psychiatry* 58, 8, 605–613.

Kuhnen, C. M. and B. Knutson (2005) "The neural basis of financial risk taking". *Neuron* 47, 5, 763–770.

Nelson, K. (2010) "Intel debuts mind-reading brain scans in NYC". *Daily News*. Retrieved from http://www.nydailynews.com/money/2010/04/08/2010-04-08_mindreading_brain_scans_debut_in_nyc.html.

Nieder, A. and E.K. Miller (2004) "A parieto-frontal network for visual numerical information in the monkey". *Proceedings of the National Academy of Sciences of the United States of America* 101, 19: 7457.

Phelps, E. A., K. J. O'Connor, W. A. Cunningham, E. S. Funayma, J. C. Gatenby, J. C. Gore, and M. R. Banaji (2000) "Performance on indirect measures of race evaluation predicts amygdala activity". *Journal of Cognitive Neuroscience* 12, 1–10.

Räikkä, J. (2010) "Brain imaging and privacy". *Neuroethics* 3, 1, 5–12.

Sententia, W. (2004) "Neuroethical considerations: cognitive liberty and converging technologies for improving human cognition". *Annals of the New York Academy of Sciences* 1013, 1, 221–228.

Sip, K. E., A. Roepstorff, W. McGregor, and C. D. Frith (2008) "Detecting deception: the scope and limits". *Trends in cognitive sciences* 12, 2, 48–53.

Zald, D. H. (2003) "The human amygdala and the emotional evaluation of sensory stimuli". *Brain Research Reviews* 41, 1, 88–123.

Zink, C. F., Y. Tong, Q. Chen, D. S. Bassett, J. L. Stein, and A. Meyer-Lindenberg (2008) "Know your place: neural processing of social hierarchy in humans". *Neuron* 58, 2, 273–283.