scripted

Volume 8, Issue 2, August 2011

RESPONSIBILITY AND THE AUTOMATICITY THREAT

Dr Tillman Vierkant[*]

Abstract

There is a common perception that brain imaging poses a great threat to our ability to control our own minds and hence to our ability to have a whole cluster of abilities (autonomy, responsibility, culpability) relevant in the context of the law. It is said that brain imaging in the future will give scientists the ability to get direct access to our inner most selves possibly even against our will. Equally, it is claimed that brain imaging might allow for mind reading and make us fully predictable, thereby rendering us helpless to thwart the predictions. In this paper I want to debunk these myths. I argue that brain imaging only seems more worrying than behavioural sciences, because it taps into a folk reductionist view of the mind according to which the mind is the brain. Secondly, I argue that predictability in the relevant sense is a myth for conceptual reasons.

Nevertheless, I think there is a real threat to our ability to control our own minds that comes from the cognitive sciences that deal with the cognitive unconscious. I end with some suggestions how this challenge can be transformed into a chance.

DOI: 10.2966/scrip.080211.184

[*] Lecturer, University of Edinburgh, UK

1. Introduction

In the Hollywood film *Minority Report*, scientists with some fancy tools are able to predict what people will do and the police can prevent crimes before they have even been committed. Potential perpetrators are then convicted for thought crimes, because even though they have not done anything physically wrong, the police know that this is only because they have been prevented from doing so by their timely arrests. Now, obviously *Minority Report* is only a Hollywood film and the predictive powers described in the film are a very long way away from reality, but the idea that scientists might be able to peek into our minds and even predict what we are going to do, before we know it ourselves, is very much part of our culture. Equally common is the worry that follows from this idea that if scientists really were to read our minds and to predict our behaviour, then that would be bad news for the notion of moral and criminal responsibility. In this paper I want to look at this worry a bit more closely.

The science that most people associate with the worry is neuroscience and in particular, brain imaging or even more precisely, fmri.[1] Going by the reports about the achievements in those sciences, especially those reported in popular science magazines or the science section of the newspaper, it is not surprising that people are worried. Looking at the actual science, scientific magazines report that neuroscientists are already able to read minds to a fascinating extent[2] and even to predict decisions before people know what they will decide.[3]

The scientists themselves always stress that they are still a long way away from true mindreading and prediction, but some of them already feel confident enough to use the rapidly growing power of these new scientific findings in commercial contexts. Lie detection is potentially a very profitable field, and as the polygraph has a very dubious reputation, one of the earliest interests from people outside science (e.g. the American military) was in brain reading technologies that could tell truth from lie more reliably. There are by now two commercial companies (No Lie Mri and Cephos) who offer commercial lie detection using fmri technology and there have been repeated attempts to get fmri lie detection into courts (successfully in India). Another example of the many ways in which neuroscience is moving out of the ivory tower of science into every day life is neuro-marketing. Traditional methods of marketing have found again and again that there is a vast gap between what people say they will buy and what they actually buy. Neuro-marketing claims that it can find out directly what people really want, and neuroscientists like Gemma Calvert and her company Neurosense have very impressive case studies[4] about how fmri has transformed advertising campaigns.

---

[1] Functional magneting resonance imaging.

[2] E Eger et al., "Deciphering Cortical Number Coding from Human Brain Activity Patterns" (2009) 19 *Current Biology* 1608-1615.

[3] C Soon et al., "Unconscious Determinants of Free Decisions in the Human Brain" (2008) 11 *Nature Neuroscience* 543-545.

[4] The case studies are available on the company's website at http://www.neurosense.com/.

2. How Real is the Challenge

2.1. Could neuroscience show that we are not at all responsible?

Let us first quickly deal with the absolute challenge. Supposing we had a perfect mindreading and prediction machine, would that not rob us of responsibility for our actions?[5] It looks like the answer to this should be yes, because if there were a machine which could make predictions about our behaviour that would always come true, no matter whether we tried to thwart them or not, then it seems we have no control over whether or not any of the behaviour will happen. It seems a very plausible assumption that we need to have at least some form of control over our behaviours to be responsible for them, but if we are unable to thwart a prediction that is made about us, then it seems that we lack control. Like Oedipus, we would not be able to avoid our fates. But is it likely that mindreading machines ever will be like that? The answer to that question must be a clear no. Independent of all the technical difficulties that stand in the way of creating a brain-reading machine that really can cope with the vast complexity of the human brain, there is one crucial problem in creating a perfect predictor. Responsibility would be impossible if absolute prediction existed, because absolute prediction renders the agent powerless to influence a certain outcome, but even a perfect mind-reader would do no such thing. As soon as we tell the subject of a prediction, we add new input to that system which the system can then use to try to confound the prediction. Even a very simple system can use new input to thwart even the most sophisticated predictor.

Now, as long as we as agents are able to carry out the same simple trick, then there is no reason to think that we are fully predictable in the sense that we would have to helplessly watch ourselves fulfilling a prediction without the power to thwart that prediction, should we want to do so. Neuroscience might be able to show that the human brain can be explained mechanistically (obviously in itself a very controversial claim), but it does not show that these mechanisms are powerless in the face of neuroscientific predictions (nor does it even attempt to do so). Thus, it is safe to say that the brain sciences will not create a predictor that will demonstrate that we are not able to control our behaviour in any way. Once we think through the above argument, it seems almost strange that one might ever have thought that this should be its aim.

However, it is important to separate what this argument does show from a seemingly similar issue that the argument has nothing to say about. This issue is determinism. According to determinism, all our decisions are predetermined. If determinism is true then we do no have what philosophers call the ability to do otherwise. Having this ability requires that before we have made a decision it is not determined which way we will decide. In order for absolute predictability to be true determinism would have to be true and it might be tempting to think that the falsity of predictability entails the falsity of determinism, but that is not the case.

The above argument is perfectly compatible with the truth of determinism. All the argument shows is that determinism does not mean that minds do not make a causal contribution to the world. A mind is able to change the world in such a way that it will normally be able to frustrate a prediction, if it knows of that prediction. But

[5] I owe this argument to Richard Holton (forthcoming).

obviously, this frustrating behaviour might well be determined. The mind might have been determined to frustrate the prediction well before it actually made the decision to do so.

So the argument does not make a weaker sense of full predictability impossible. Given that the decision of the mind might be predetermined, it might still be possible to predict what a mind will do with 100% accuracy. All that the argument has shown is that it is impossible to make such predictions if the subject has been made aware of them and has a motivation to frustrate them.

But one might think that this weaker sense of predictability is enough to make human responsibility impossible, and that if neuroscience shows that this form of predictability is possible, then that would be enough to show that there is no responsibility.

Perhaps surprisingly, the majority of philosophers who think about the concept of responsibility are of the opinion that determinism and responsibility can coexist. One standard argument for this is that what we really mean when we say that we are responsible for our actions is not that we could have acted differently if we evaluated the situation in exactly the same way as we did when she made our decision, but that we could have acted differently, if we had evaluated our reasons to act differently.

If that is true, then what makes us responsible for our actions is that we can understand reasons and that our actions depend on whether we think we have sufficient reason to act.

Ultimately however, whether or not this is the right way to understand the conditions for human responsibility does not really matter for this paper. Here we are interested in the question of whether the advance of neuroscience provides new arguments that show that we are not really responsible for our actions. The above is only important here because it shows that this is unlikely. This is because it is unclear what exactly the necessary conditions for responsibility are and whether they are - or are not - compatible with the truth of determinism. These are clearly conceptual issues that cannot be decided by neuroscientific experiments.

Neuroscience as an empirical science mainly relies on a working assumption of the truth of determinism. In this sense, it does not provide evidence for or against the doctrine. It has even less to say about the question of whether determinism is compatible with responsibility.

We can safely conclude that the advances of neuroscience will not show that we are not responsible for our actions. They will not show that our behaviour is not under our control because it is predetermined. This is because the weaker reading of having no control, which is implied by determinism, is not strong enough to make responsibility impossible. Even though a strong reading of having no control would imply the loss of responsibility for one's behaviour, this does not matter, because we have good conceptual reasons to believe that the strong reading is simply false.

2.2. What is the challenge then?

In the last section it was argued that neuroscience does not and will not show human responsibility to be an incoherent idea by reading and predicting minds and thereby showing that we do not control our behaviour. It was shown that predictability would only be a threat to responsibility if it were understood in such a strong way that the

doctrine seems very likely to be false. But predictability is not the only threat that the advancing neurosciences have in store for us. In this section I want to discuss a slightly less direct threat that can be intuitively constructed from the ability of the neurosciences to get direct access to our brains and read our minds. I will argue that the challenge is real, but that the focus on fmri is a red herring.

We live in a society that makes it very difficult to hide completely away from the piercing public gaze (CCTV, tracking of mobile phones, Internet, data files in government agencies and so on), but we still feel that at least our thoughts are private. Only we have direct access to them and as long as we keep them to ourselves, nobody else has. This also means that it is very difficult for others to control what we think. Thoughts are free, as the German proverb says, because nobody but us knows what we are thinking. One way to think about the challenge to this freedom is that neuroscience endangers the special private relationship that people have with their own minds and thereby endangers the special freedom from external control that our minds are thought to have. According to a broadly (and slightly caricatured) Cartesian line of thinking, we might be wrong about the world, but our minds are fully transparent to ourselves. All other people can only use indirect behavioural measures to find out what is going on in our mind, but we only need to be introspective to know what is going on. Cartesian substance dualism is not very popular nowadays, but the idea of the mind as somehow hidden away from the world is still very powerful. What takes up the place of the Cartesian mind in our materialistic-minded times is the brain. The brain is the place where our minds are stored, away from the public gaze and safely behind the firm walls of the skull.

The closer we are to such a picture of the mind, the more understandable it becomes that the brain sciences are perceived as a qualitatively bigger threat to the privacy of our minds than all other sciences combined. All behavioural sciences only get indirect measures of the mind, but the brain sciences, as it were, can directly look into our innermost sanctuary. They can directly observe the workings of the mechanisms that support the conscious self. Our innermost thoughts can now be directly examined.

If one believes in such a picture of the mind that might as well explain a fascinating set of recent data. There is very good evidence that somehow neuroscience and in particular fmri does not only impress us, because of its findings and scientific conclusions from them, but because the brain images themselves have an enormous psychological power.[6] Weisberg and colleagues found that people are much more likely to accept faulty psychological arguments, if they were backed up by irrelevant brain pictures. This does not seem particularly surprising if it is true that we simply identify the mind with the brain. In this case, seeing the brain is like seeing the mind at work and it is not surprising that the intuition of such a direct access to what has been completely inaccessible throughout the history of mankind might impress us more than it should.

But obviously, if the threat from neuroscience were dependent on such a strong Cartesian view of the mind, then it would be quite blunt, because we have all known, since the days of Freud at the very least, that the mind is not fully transparent to itself.

---

[6] D Weisberg et al., "The Seductive Allure of Neuroscience" (2008) 20 *Journal of Cognitive Neuroscience* 470-477.

Until recently, the unconscious was at least thought to be only the, as it were, irrational dark side of our psyche, while all the hard cognitive work was supposed to happen in the conscious part of the mind. Social psychology has taught us that even this is far less obvious than we might have thought.[7] There is now an amazing wealth of findings about the so-called "cognitive unconscious"[8] that seems to demonstrate that most of the higher mental processes that had been associated with the conscious mind can function entirely in the absence of conscious awareness.

But does the fact that our mind does not only contain the conscious self lead to a blunting of the neuroscientific challenge to the privacy of our minds? It seems not. On the contrary, if this is how we should think about our minds, then one might think neuroscience has now become even more threatening, because now it is not only able to nudge closer and closer to our privileged introspective knowledge about ourselves, but it might actually find out more about our minds than we can ever know from the first person perspective by introspecting. Neuroscience can know what is really happening in our brains and therefore in our minds.

Neuroscience can reveal our true preferences, not what we, often falsely, take our preferences to be from the first person perspective. As neuro-marketing and the interest of the military shows, there are many people who do not only want this knowledge for purely scientific reasons, but in order to manipulate our minds. This looks like quite a significant threat. Neuroscience might not show that we cannot be responsible agents, but it can help others to undermine our control over our behaviour, by giving them the knowledge of how to manipulate our minds without our knowledge.

2.3. Is the intuitive fear of the power of the neurosciences justified?

The neurosciences seem threatening because they seem to have a special and qualitatively different way of finding out about the content of our minds, but is it really the case that their access is privileged? One important point that has to be made here is purely practical. Even if one were to buy into the different quality of neuroscientific knowledge about our minds, then it would still be the case that – at least at the moment – the abilities of the neurosciences to find out private thoughts pale in comparison to the level of surveillance that can be achieved by traditional methods.[9]

But here I want to examine a more theoretical worry about the importance of neuroscience for understanding the mind. In the last section I claimed that the wealth of recent findings on the power of the cognitive unconscious pose a new and significant threat to our understanding of how our minds are controlled. I claimed that understanding the cognitive unconscious has taught us that our minds are far less transparent to us than we ordinarily think and that much of what we do is done without us knowing the true mental causes for our behaviour.[10] This I claimed meant

---

[7] T Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Cambridge, MA: Belknap Press, 2002).

[8] J Kihlstrom, "The Cognitive Unconscious" (1987) 237 *Science* 1445-1452.

[9] N Levy, *Neuroethics: Challenges for the 21st century* (Cambridge: CUP, 2007).

[10] We call this the "zombie challenge" in a forthcoming collection. T Vierkant (forthcoming).

that we have far less control over attempts to manipulate our minds, because we simply might not notice the manipulation. However, what is not clear these claims are how important the role of neuroscience is for the challenge.

Here, it might turn out that the understanding of the mind as not only conscious, which seemed to make the challenge even stronger, in actual fact significantly undermines it. As long as understanding minds is about understanding the conscious self, which was supposed to be somewhere in the brain, there is an obvious qualitative difference between what neuroscience does and what the other cognitive sciences do. Only neuroscience can directly observe the conscious self. Once we understand the importance of the vast array of unconscious subsystems for even some of our high cognitive achievements, the ability to directly look into the brain becomes less special. It is now far less interesting to be able to access the one special place where the mind, as it were, resides. What matters now, is simply to observe what the many systems do. The many unconscious systems that make up so much of our minds are not little homunculi sitting in our brains, but rather they are mechanisms that provide functionalities which when combined explain our overt behaviour. In order to study these functions it becomes less of a priority to look, as it were, inside the mechanism itself to be able to observe a first person like mental theatre. Instead, what matters most is simply the rules according to which these mechanisms produce output. Even more importantly, these mechanisms are often not fixed structures in the brain, but highly flexible, and highly integrated in the environment. Simply looking at the mechanism of the brain will not be enough to tell us what it is that the system – of which the brain is only a part – is doing.

If this is how we should think about the mind, then it is still true that it is quite possible for third parties to find out what is going on in our minds even if we do not know ourselves, but we should not expect the neurosciences to be especially privileged in doing so. This is because there is not one single place in the brain or anywhere else where one can as it were observe the theatre of the mind. Rather our minds are deeply embedded in the environments they interact with. Because of this, studying behaviour should be at least as good a route to understanding what the mind is doing as looking at the brain.

In fact, this is what we find at the moment. Many of the most fascinating and counterintuitive lessons about how our minds work do not come from fmri studies, but from the laboratories of social psychologists. In fact, the very term "cognitive unconscious" is a term coined in psychology. The Neuro-marketing mentioned in the introduction, which seemed so threatening because it allows the manipulation of our minds without our knowledge, actually uses many techniques which come from social psychology research.[11] Even the success rate of fmri lie detection is not really much better than the traditional polygraph and has been not widely tested with subjects who have a serious interest in not cooperating and who try to confound the machine.[12] In conclusion then, as we are not Cartesian creatures with an inner sanctuary that is safe from direct manipulations, it is quite possible to manipulate our minds without us

---

[11] See e.g. Tony Greenwald's 1998 work on the IAT: A Greenwald, D McGhee, and J Schwartz, "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test" (1998) 74 *Journal of Personality and Social Psychology* 1464-1480 and J Bargh and T Chartrand, "The Unbearable Automaticity of Being" (1999) 54 *American Psychologist* 462-479.

[12] See, e.g. N Levy, at note 9 above.

knowing, but the tools to do this do not have to be neuroscientific, but can as easily come from the behavioural sciences. Focussing on neuroscience in this case almost seems to serve as a distraction from what should really concern us.

3. So is the Cognitive Unconscious Threat Very Serious then?

There is no denying that understanding the cognitive unconscious is an enormously powerful tool for manipulating minds, but obviously it cannot only be used for sinister means. The cognitive unconscious revolution together with many other strands of modern cognitive science has taught us what enormously situated creatures we are. Every new account of human autonomy will have to take that into account.[13] An autonomous creature on such a picture should not believe that we can live our lives uninfluenced by the environments that we are in, and needs to know that we will not always know introspectively which way our environments are influencing us.

One fascinating and provocative idea of how this knowledge could be put to concrete use in the context of the law is in the work of Professor Susan Hurley. Hurley argues in her paper "Imitation, Media Violence and Freedom of Speech"[14] that the distributors of violent media entertainment should be at least partially responsible, if there were (as she argues there is) an increase in real violence as a result of increased exposure to media violence. She argues that this is the case, because the link between media and real violence functions via an automatic imitative link which individuals do not control and are not aware of. The conclusions that Hurley draws are surely very controversial, but she shows the way that the cognitive sciences can help us to re-examine questions of responsibility with a fresh perspective which is enriched and not threatened by the progress of the sciences of the mind. The sciences help us to understand how situated our cognition is and once we are aware of that, we can then begin to think about how to account for that in our understanding of sharing responsibility between the individual and her environment and especially the society and communities that form the environment.

Once we become aware of the limitations of individual responsibility, the threat of situatedness can become a chance for autonomous self control. We can then protect ourselves against manipulations by others and, even better, consciously shape our environments to manipulate our own psychology according to our values.

---

[13] For attempts using the knowledge about the situatedness of our cognition for a reconstruction of the concept of responsibility see e.g. T Vierkant, note 10 above.

[14] S Hurley, "Imitation, Media Violence, And Freedom of Speech" (2004) 117 *Philosophical Studies* 165-218.