International Journal of Law and Psychiatry 35 (2012) 104–111

Contents lists available at SciVerse ScienceDirect

International Journal of Law and Psychiatry

Brains in context in the neurolaw debate: The examples of free will and "dangerous" brains

Stephan Schleim *

Theory and History of Psychology, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

ARTICLE INFO

Available online 29 January 2012

Keywords:
Neurolaw
Neuroethics
Free will
Legal responsibility
Ventromedial prefrontal cortex
Phineas Gage

ABSTRACT

Will neuroscience revolutionize forensic practice and our legal institutions? In the debate about the legal implications of brain research, free will and the neural bases of antisocial or criminal behavior are of central importance. By analyzing frequently quoted examples for the unconscious determinants of behavior and antisocial personality changes caused by brain lesions in a wider psychological and social context, the paper argues for a cautious middle position: Evidence for an impending normative "neuro-revolution" is scarce and neuroscience may instead gradually improve legal practice in the long run, particularly where normative questions directly pertain to brain-related questions. In the conclusion the paper raises concerns that applying neuroscience methods about an individual's responsibility or dangerousness is premature at the present time and carries serious individual and societal risks. Putting findings from brain research in wider contexts renders them empirically investigable in a way that does not neglect psychological and social aspects of human mind and behavior.

1. Introduction

Advances in neuroscientific knowledge and technology play an important role in recent debates on the interaction between science and society. Some have perceived neuroscience to pose a challenge to our notions of free will and responsibility, thus potentially threatening established societal practices concerning accountability and punishment for one's deeds (Greene & Cohen, 2004; Sasso, 2009; Sie & Wouters, 2010). For example, Greene and Cohen argued that after intuitions of the public at large about responsibility and punishment have been changed through neuroscience, the public will not accept contradicting norms and require legal doctrines to be changed accordingly (Greene & Cohen, 2004). Others consider possible applications such as in forensic contexts in order to improve current practices, for example, assessing the credibility of testimony or the dangerousness of people, without the necessity of overthrowing their normative basis (Aggarwal, 2009; Busey & Loftus, 2007; Goodenough & Tucker, 2010; Vincent, 2011). Indeed, there are recent cases where neurogenetics and neuroimaging evidence led to mitigated sentences because they putatively demonstrated a tendency towards aggressive behavior or the presence of a mental disorder.[1]

The societal impact of such neuroscience knowledge or technology innovations depends on neuroscientists' capability to convince people of their methods' possibilities. Psychological research indicates that adding neuroscientific knowledge or neuroimages to otherwise identical psychological explanations moderately increases their perceived scientific credibility (Keehner, Mayberry, & Fischer, 2011; McCabe & Castel, 2008; Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). Analyses of media coverage of functional magnetic resonance imaging (fMRI) research by Racine and colleagues or Caulfield and colleagues shows that such reports are predominantly optimistic and frequently suggest the reduction and/or identification of psychological and societal questions with brain-related questions (Caulfield, 2004; Caulfield & Ogbogu, 2008; Racine, Bar-Ilan, & Illes, 2005; 2006; Racine, Waldman, Rosenberg, & Illes, 2010). Indeed, when writing about the future of his discipline, Nobel laureate Roger Sperry predicted that "philosophies, religious doctrines, world-models, value systems, and the like will stand or fall depending on the kinds of answers that brain research eventually reveals" and stated that it "all comes together in the brain" (Sperry, 1981, p. 4). Yet, others argue that neuroscience technologies such as fMRI are not ready to be translated into individual practical applications due to methodological and conceptual limitations (Logothetis, 2008; Schleim & Roiser, 2009).

In this article I analyze two popular domains which are of central importance to neurolaw: the first is related to theoretical neuroscience knowledge and how it is understood psychologically in the context of free will, the second concerns the practical technological application to identify "dangerous brains" in forensic settings. Taking historical as well as recent scientific cases seriously, I argue that

* Tel.: +31 50 363 6244; fax: +31 50 363 6304.
E-mail address: s.schleim@rug.nl.

[1] See these media reports on two cases decided in Italy in 2009 and 2011 widely discussed in the scientific community, http://www.nature.com/news/2009/091030/full/news.2009.1050.html and http://blogs.nature.com/news/2011/09/italian_court_reduces_murder_s.html, both accessed on 10/4/2011.

0160-2527/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijlp.2012.01.001

neuroscience explanations at least partially, yet essentially, depend on social as well as psychological context knowledge. In the light of this context knowledge, neuroscientific innovations can be understood better as posing scientific challenges instead of causing social revolutions. Accordingly, putting findings from brain research in wider contexts allows an understanding of them that does not contradict but instead support and complement present views on what it means to be human and how to make forensic assessments. I conclude that instead of overthrowing established social practices, particularly in the legal context, neuroscience most likely will gradually improve it in some cases and turn out to be less relevant than perceived by some in others, at least given the present state of knowledge.

2. Free will in context

In 1917, the then leading psychologist Sigmund Freud published an essay in which he describes three humiliations that mankind suffered from science: first, the cosmological humiliation, based on Copernicus's finding that the earth is not in the center of the universe; second, the evolutionary one, implied in Darwin's theory with the consequence that humans and apes are biological relatives; and third, the psychological humiliation, posed by Freud's own psychodynamic theory holding that humans only have limited control of their sexual drives and limited conscious access to their own psychological processes. He summarized the last one in the statement that the ego is not the master in its own house (Freud, 1917).

This idea, although now almost 100 years old, has become remarkably popular in recent neuroscience research. Already in the 1980s, Benjamin Libet's famous experiments on the neural processes of consciousness with the finding that the conscious intention to press a button is temporally preceded by a readiness potential recorded from the premotor cortex by a few hundred milliseconds (Libet, Gleason, Wright, & Pearl, 1983) was interpreted in the same manner, namely, that conscious events are not the causes of our actions, and is still discussed in this respect nowadays (Aharoni, Funk, Sinnott-Armstrong, & Gazzaniga, 2008). Although the experiment has been criticized in many respects, neuroscientifically, psychologically, and philosophically (Árnason, 2011; Haggard, 2008; Roskies, 2010), I find it noteworthy that Libet already reported that the readiness potential could even be measured in the absence of the action, when the subject chose to have a "veto" and not to press the button after the conscious intention occurred, as explicitly allowed by the experimental instruction (Libet, 1985). Regarding the simple causal model of temporal subsequence, the readiness potential could thus not be perceived as a strict cause of the action, since the former could also occur without the latter. The finding that the readiness potential does not differ between decisions to move and decisions not to move has been reproduced recently and is interpreted as evidence against unconscious movement initiation (Trevena & Miller, 2010).

Nevertheless, a recent study that was designed as a variant of Libet's original experiment, though carried out with some technical changes due to the setting in a modern fMRI scanner, where the subjects could choose between two different button presses, left and right index finger, but without being allowed to have a "veto", is interpreted as uncovering unconscious determinants of our actions in brain activity as early as ten seconds before the intention becomes conscious (Soon, Brass, Heinze, & Haynes, 2008). As with Libet's experiment before, inferences on free will, responsibility and the legal system are drawn from these findings and communicated to an interdisciplinary audience and the public (Smith, 2011; Welberg, 2008). The upshot of the argument is not generally that determinants of human behavior are found, which philosophers advocating the so-called stance of compatibilism consider compatible with or even a prerequisite of free will (Árnason, 2011; Roskies, 2006; Walter, 2001), but that behavior is determined by processes of which we have no control, as Freud already claimed, because they are essentially unconscious.

Although the experiment by Soon and colleagues can be criticized on various grounds, for example, because 60% predictive value of the side where the subjects will press is not extraordinarily high, since 50% is already chance level, or that two thirds of the subjects had to be excluded from the experiment after a behavioral pre-test or even after fMRI data acquisition, because they were too slow or had a strong tendency towards one side in their responses, I would like to discuss a conceptual as well as a neuroscientific point regarding their interpretation instead.

First of all, what did the subjects have to do? According to the paper, the instruction was: "At some point, when they felt the urge to do so, they were to freely decide between one of two buttons, operated by the left and right index fingers, and press it immediately" (Soon et al., 2008, p. 543). In the supplementary material, though, the instruction runs differently, namely: "Subjects were told to relax and to press either the left or right button with the index finger of the corresponding hand immediately when they became aware of the urge to do so" (Soon et al., 2008, p. S15). According to the first version, the cognitive model assumes an urge to make a decision, a decision, and a button press; but the second one does not refer to any decision process at all and is restricted to an urge and a button press instead. The conceptual problem of understanding what it means to have the urge to press a button, an experience one rarely has in everyday life, was already prevalent in Libet's original experiment, as he used the same kind of instruction (Libet, 1985). Since the subjects are asked not to plan their decisions previously and to react as spontaneously and fast as possible, yet to leave some time between their decisions, one might wonder whether they actually tried to simulate a kind of binary random number generator. Their behavioral responses indeed resemble random behavior (Soon et al., 2008). This undermines the psychological meaning of the kind of decision the subjects had to make, if it was a decision in a rich sense at all (Hartmann, 2004).

My second kind of critique concerns the brain areas from which the experimenters could calculate the prediction up to ten seconds before, that is, the lateral frontopolar cortex and the precuneus/posterior cingulate cortex. If the genuine determinants of the subjects' decisions were unconscious, one would expect brain activation in areas frequently related to unconscious biases, such as the limbic system or basal ganglia, to allow the prediction long before the intention to make the decision becomes conscious. However, this is precisely what the researchers have not found, but instead they themselves refer to these brain areas as a network of high-level control areas (Soon et al., 2008) and indeed lateral areas in the prefrontal cortex are frequently associated with reasoning processes, controlled decisions, and explicit rule application (Bunge, 2004; Miller & Cohen, 2001; Schleim, Spranger, Erk, & Walter, 2011; Siegel & Douard, 2011) and the precuneus is one of the central areas related to consciousness (Laureys, Owen, & Schiff, 2004).

How can the researchers then know that the determinants they discovered are unconscious? In a trivial sense, all brain processes are unconscious, since we ascribe "conscious" to psychological processes people are aware of; and if we had a definition of consciousness in neural terms, the prefrontal cortex as well as the precuneus would play an important role. In the study, subjects were asked to assign the time of the conscious intention to perform the action to a certain moment in a forced-choice situation. This model of consciousness presumes that a psychological process is unconscious and suddenly "jumps" into consciousness; it does not allow the possibility that a process gradually enters consciousness and certainly the subjects were conscious and had conscious processes occurring during the whole period of the experiment. Besides the limited forced-choice self-report, the experimenters have no source of information whether the brain activation they calculate the prediction from is related to