International Journal of Law and Psychiatry 65 (2019) 101344

Contents lists available at ScienceDirect

International Journal of Law and Psychiatry

In defense of free will: Neuroscience and criminal responsibility

Paul G. Nestor

Department of Psychology, University of Massachusetts Boston, United States
Department of Psychiatry, Laboratory of Neuroscience, Harvard Medical School, United States

ARTICLE INFO

Article history:
Received 23 January 2018
Accepted 5 April 2018
Available online 22 April 2018

ABSTRACT

Is neuroscience the death of free will and if so, does this mean the imminent demise of the psycho-legal practices related to insanity and criminal responsibility? For many scholars of neuro-jurisprudence, recent advances in brain sciences suggesting that the perception of free will is merely illusory, an epiphenomenon of unconscious brain activity, do indeed undermine our traditional understandings of moral and legal responsibility. In this paper, however, we reject this radical claim and argue that neuroscientific evidence can indeed reveal how free will actually works and how its underlying neural and perceptual machinery gives rise to our sense of responsibility for our actions. First, the experience of free will is recast in terms of neuroscientific studies of agency and willed action. Second, evidence is presented of a neural network model linking agency to widely-distributed brain areas encompassing frontal motor and parietal monitoring sites. We then apply these findings to criminal responsibility practices by demonstrating (a) how the experience of intentionality and agency is generated by specific interactions of this discrete frontal-parietal network, (b) how mental disease/defect may compromise this network, and (c) how such pathologies may lead to disturbances in the sense of agency that often are central to the phenomenological experience of psychosis. The paper concludes by examining criminal responsibility practices through the lens of cultural evolution of fairness and cooperation.

Contents

1. Introduction

"Neuroscience in the Courts — A Revolution in Justice?" reads the headline of a 28 July 2006 piece appearing in *Science* by Earl Lane. The *Science* article reports on a seminar organized by the American Association of the Advancement of Science (AAAS) for state and federal judges who convened to learn from experts about the current and future prospects of neuroscience research and its implications for jurisprudence. This seminar is just one example of the growing scholarly interest that has unfolded over the past 20 years in seeking to apply the advances in the burgeoning brain sciences to law in general, and criminal law, in particular (Zeki & Goodenough, 2004). Such scholarship resonates with excitement and promise as it envisions a future of jurisprudence in which our understanding and judgment of such vexatious problems of honesty, truthfulness, punishment, and justice can be advanced, enlightened, if not solely determined by neuroscience (see Gazzaniga, 2005, 2011; Greene & Cohen, 2004; O'Hara, 2004; Zeki & Goodenough, 2006). From this vision emerges "neuro-jurisprudence" as a transformative force that will in time reshape and recast the law — a veritable revolution as headlined!

https://doi.org/10.1016/j.ijlp.2018.04.004
0160-2527/

For the legal system, however, a future dominated by neuroscience raises serious philosophical and practical questions. Chief among these is the influence that advances in neuroscience will have on matters of criminal responsibility, and the practice of forensic psychiatry and psychology. Indeed for criminal justice, there is little question about free will as the vital source of human action, as sacrosanct assumption as it has been for many religions throughout history. Yet, in stark contrast stand many investigations emanating from neuroscience and philosophy of mind that forcefully argue against, if not downright reject the causal efficacy of free will in decision making and voluntary action (e.g., Gallagher, 2006; Wegner, 2002). For these researchers, free will, as traditionally understood, represents an antiquated concept — a remnant of a dualistic folk psychology rendered false by modern brain science (Churchland, 1990; Crick, 1994). And so as Greene and Cohen (2006) forecast, as the legal system increasingly comes to understand the causal inefficacious and therefore true illusory nature of free will, the doctrine of criminal responsibility will be upended because "in a very real sense we are all puppets. The combined effects of gene and environment determine all of our actions." (p. 217).

This paper traces the origins of this provocative assertion of free will as an illusion across philosophical, psychological and neuroscientific literatures. The evidentiary basis for this claim is critiqued, and borrowing from Mark Twain, reports of the death of free will are considered to be greatly exaggerated. Recent neuroscience findings related to decision making and willed action are then reviewed. These studies provide the basis for what I propose as an inchoate neural model of conscious intention, the aim of which is to help illuminate the scientific bases for willed action that is central to the legal doctrine of criminal responsibility. Last, the paper concludes by proposing that the scientific and ecological validity of free will can be best understood as a product of group selection in the evolution of cooperation and fairness, which in turn can be viewed as a conceptual framework for viewing why free will exists and how it functions in legal matters of criminal responsibility and insanity.

2. Why is free will a problem?

For philosophers free will can be a problem because of its dualistic baggage. For psychologists, free will can be a problem because of its apparent illusory nature as a causal agent in driving human actions. Free will does not cause specific acts in real time, that is, in the instant in which they happen, but rather it follows the decision to act (Vohs, 2010). The illusion lies not in the experience itself but in the faulty attribution of it as the cause of action. For neuroscientists, free will is a problem because as will be discussed below, neurophysiological studies have offered evidence that electrical potential precedes intentional movement by around −550 to −250 ms before a subject is aware of the decision to act. For neuroscientists, these data suggest that the experience of human agency is an epiphenomenon that can be simply understood as the product of brain machinations that implement it. For neuroscientists, the prevailing monism of science requires that brain activation and subjective experience be essentially one in the same, so tightly and irrevocably coupled, in time and space, to form a single, unified reality. Thus, the experience of conscious free will must have a corresponding, time-locked, simultaneous brain state. In the absence of such, then the subjective experience of free will can be dismissed as an illusion.

3. Philosophy

The philosophical question of free will is often framed dialectically: we are either agents of free will or agents of determinism. For many philosophers, determinism precludes free will, as Roskies (2010) wrote, "If there is only one course the universe could take, and we are part of that universe, we could not do other than we do, and thus we do not have real choices or options, and so no free will." (p. 154). The root of this philosophical problem is how to accept free will without appealing to a dualism of mind and body that conceives of the brain as the repository of desires and intentions emanating from a nonphysical entity, such as a soul. Thus, for many researchers, to accept the idea of free will means rejection of the prevailing philosophy of science, monism in which physical events and human actions are determined by lawful, mechanistic relationships among interacting material substances (Horst, 2011; Roskies, 2006, 2010).

However, as both Horst (2011) and Roskies (2010) proposed, a mechanistic view of human actions governed by lawful interactions does not inevitably lead to a rejection of free will. For example, Horst (2011) offered a different and nuanced perspective on the "free will problem" for the monism of neuroscience. He argued that monism has too often erroneously assumed that a commitment to scientific laws necessitates a commitment to determinism, and thus a denial of free will. Rather he made the compelling case that physical states and natural laws are more probabilistic than deterministic, more akin to empirical generalizations than to hard-and-fast, strict, universal rules. Horst showed that not only do human sciences have "ceteris paribus" ("other things being equal") laws, but so too do the natural sciences. This view of scientific laws does not, of course, constitute proof that the will is free. But as Horst asserted, by the same token, "it is a proof that free will is not inconsistent with a commitment to scientific laws, properly understood." (p. 119). Or as Roskies (2010) noted, the scientific validity of free will does not hinge on the question of determinism. Moreover, Roskies argues that the question of determinism will never be answered by neuroscience, but rather is ultimately a question for physics and quantum mechanics (see also Searle, 2010). Thus, the important point of these philosophical analyses is that neuroscience can elucidate the brain bases of free will and responsibility once these concepts are considered independently of the question of determinism. Therefore decoupling ceteris paribus laws from determinism may very well lay the conceptual ground work for monism to study free will as an open scientific construct.

4. Psychology

Psychological studies have also addressed the problem of free will. These studies have raised the interesting and intriguing question as to whether free will is real or merely an illusion. Among the most influential work in this area is the research of Wegner (2002) who in a masterful monograph based on a review of experimental psychology literature concluded that the conscious experience of will as causing action is an illusion. Following several other theories (e.g., Brown, 1989; Langer, 1975; Libet, Gleason, Wright, & Pearl, 1983; Spence, 1996), Wegner's thesis rejected the ontological reality of free will as the mental engine of causation. From his perspective, the experience of free will is a product of the mind's propensity to infer erroneous casual connections on the basis of association or correlation stemming from the co-occurrence of thoughts and actions. Free will, Wegner argues, is rooted in this proclivity for inferring apparent mental causation for related but independent events or processes. Thus as Wegner (2002) wrote, "This means that people experience conscious will quite independently of any actual causal connection between their thoughts and their action." (p. 64).

A key evidentiary source of support for Wegner's theory is experimental work showing that research participants can be duped into taking responsibility for actions over which they had no control. For example, Wegner and Wheatley (1999) devised an experiment to investigate whether college students would come to believe that they had moved a cursor if they had to simply anticipate where the pointer would go in advance of its actual movement. Each research participant, paired with a study confederate, sat facing each other at a small table on which was a 12-centimeter-square board mounted on a computer mouse. Participants were instructed to keep their fingers on the side of the board (akin to Ouija board game) closest to them, and the

confederate did likewise for the side closest to him/her. A computer screen visible to both participant and confederate displayed about 50 small objects and the cursor could be moved to these different spots on the screen. Participants listened to music and words throughout the experiment, which included the names of the objects on the screen that were used to evoke thoughts. The idea of the experiment was to create in the research participants the illusion of control of moving the cursor to different objects on the screen (e.g., swan), even though it was the confederate who produced the movement. Their results showed that the closer in time participant heard the name of an object (e.g., swan) and the cursor landing on that object, the stronger their misperception was that they had deliberately moved the cursor to that location. Wegner (2002) interpreted these findings as suggesting that the illusion of intentionality can be induced by the frequent co-occurrence of thoughts (e.g., heard word) followed by actions (e.g. cursor location).

The research of Wegner and others show the illusion of willed action can be induced experimentally. And perhaps like other well-known systematic biases of the human mind (e.g., Tversky & Kahneman, 1974), we can mistakenly perceive causality between our thoughts and actions. And like other universal blind spots or mental tunnels of our minds (Piatelli-Palmarini, 1994), the illusory quality of the perception of free will strike us as so counterintuitive to our subjective experience and it elicits such a strong sense of disbelief and resistance when brought to light. It is unknown whether these strong reactions reflect an important adaptive function that free will may have played in our evolutionary history in promoting fitness, morality, and cooperation (Gingerenzer, 2008; Haidt, 2007; Nowak & Highfield, 2011; Wilson, 2002). By the same token, however, Wegner's work showed that people can make mistakes about the purview of their volitional actions. This does not mean that the perception of free will as causing actions is merely illusory (Roskies, 2010). Under certain conditions, our senses are subject to illusions, but this does not mean that our senses are not valid coding for information. Moreover the experimental tasks used in these psychological studies are hardly representative of the kind of decisions for which questions about free will really matter, and they certainly bear no relevance to real live issues of criminal responsibility.

5. Neuroscience

The seminal work of neurophysiologist Benjamin Libet (e.g., Libet, 1985) is often cited as a primary source of neuroscientific evidence for debunking the notion of free will as the causative agent of action (Donald, 2010). Libet's experiments addressed the timing of brain signals in relation to simple motor decisions and perceptual awareness. On Libet's self-paced decision-making task, participants are free to press a button whenever they chose to do so. Libet only asked that his participants note on a clock that was in front of them when they had decided to press the button. He recorded electrical brain activity before, during and after participants decided whether to press the button. Libet's results showed that neuronal firings in a particular region of the brain located in the frontal lobe known as the supplementary motor area preceded the conscious choice to press the button. These neural recordings suggested that this selective brain electrical activity, known as the "readiness potential" came close to a full half second before people had made their intentions known by reading the position of a moving second hand on a clock face. The Libet finding raised the intriguing possibility that we unconsciously decide, as reflected in the readiness brain potential, well before we think we have consciously chosen to act; that is, having made up our minds might be viewed as a result of neural events, rather than as the cause.

There are, however, several important limitations to these investigations. First, the readiness potential of the pre-supplementary area reflects just that, a well-established brain wave that is associated with generalized preparation (Walter, Cooper, Aldrige, McCallum, & Winter, 1964), which has been shown to precede the initiation of any voluntary action (Donald, 2010). As Donald (2010) noted, Libet's results demonstrated only that the brain anticipated the simple decision to move a finger long before the actual decision was carried out. The second limitation pertains to the neurophysiology of the readiness potential. This brain wave represents slow-acting, low-frequency electrical oscillations, which more likely reflect neural mechanisms of anticipation, preparing, and planning of a response rather than the actual implementation of a movement (Donald, 2010) The latter voluntary action of movement is reflected, neurophysiologically, as a "command potential" that has been shown to have a different time course, morphology, and cortical source than the readiness potential (Donald, 2010). Extrapolating from more recent neurophysiological data (Canolty et al., 2006) for the Libet experiments, then the slow-moving readiness potential may serve as a tuning function to prepare the brain for action. If this is the case, the claim that such electrical activity negates free will is much more difficult to support on the basis of neurophysiology.

In fact, a recent neurophysiological study by Rigoni and colleagues (Rigoni, Kuhn, Sartori, & Brass, 2011) investigated the interesting question as to whether inducing disbelief in free will in people would alter their readiness potential corresponding to voluntary motor preparation. These researchers recorded event-related potentials while participants performed the Libet task. The event-related potential (ERP) methodology allowed the researchers to examine the time course of the slow negative-going wave of the readiness potential. The ERP analysis distinguished two temporal components: an early component readiness potential occurring up to 2000 ms before movement execution, reflecting motor preparation and presumed to be generated by the pre-supplementary motor area, and a late component readiness potential occurring around 500 ms prior to movement onset, reflecting the specific programming for movement execution, and presumed to be generated by primary motor cortex and supplementary motor area. Rigoni and colleagues sampled ERP activity across 11 200-ms windows covering the epoch from −2300 to −100 ms before the participant made a button press. Their results indicated that in relation to controls, participants who were induced to disbelief in free will showed reduced ERP amplitude in the readiness potential. In addition, their results revealed a significant interaction of time (11 time windows) and group (control, no free will), and with follow-up analyses linking induced disbelief in free will with reduced amplitude of the early but not late readiness potential.

6. Neuroscience of willed action

Most laboratory experiments of voluntary action pose considerable challenges for research design and logic (Haggard, 2009). The design or methodological problem lies in the nature of the self-paced movement task that is often used to measure voluntary action. Here a research participant can choose what movement to make, when to make it, or whether to make it at all. It is after all a task designed to measure volition. And the validity of the experimental task hinges in large part on the extent to which it captures volition. But as a putative measure of an exclusively, internally-generated action that is a product of free will, it cannot and, by definition, should not provide the kind of precision in input and output controls that are typical of laboratory-based experimental tasks. As Haggard (2009) noted, this inherent lack of experimental control can seem unsatisfactory from a research design perspective. He further pointed out the paradox of experimental studies of volition in which the task instructions can be construed as a command to "Have free will-now! (p. 732).

Recently Desmurget et al. (2009) addressed these limitations by investigating willed action using a rather powerful research design in which direct electrical stimulation was delivered to specific brain sites. For this study, the research participants were fully awake patients with brain tumors who were immobilized in a stereotactic frame as part of preparations for neurosurgery. Four patients had tumors located anterior to the central sulcus (pre-central tumors) and three patients

had tumors located posterior to the central sulcus (post-central tumors) With this methodology, the input of direct electrical stimulation could be precisely calibrated, allowing for a degree of experimental control that has been noticeably absent in prior studies of willed action. An additional critical advantage of the design was that the neural underpinnings of conscious experience could be directly tested. That is, stimulation could be delivered to specific exposed areas of the brain, such as the pre-supplementary motor area of the frontal lobe that has been hypothesized to be causally linked to the distinct conscious experience of intentionality that accompanies willed action. And indeed the results of the Desmurget et al. study shed light on how the brain may produce the experience of intentionality.

First, their data suggested that not only the frontal cortex but also the parietal cortex may be involved in the conscious experience of intention that accompanies willed action. Second, and perhaps most important, the post central tumor patients, who received electrical stimulation of inferior parietal sites, experienced a clear intention and desire to move, without any overt movement as recorded by EMG. Most striking was that higher intensities of the electrical stimulation produced in these patients the illusion of movement, though no movement had really occurred. For example, Desmurget and colleagues described the self-report of one patient after low intensity stimulation of the inferior parietal areas, as saying "I felt a desire to lick my lips" and with increased electrical stimulation, saying "I moved my mouth, I talked, what did I say?" (p. 812). Desmurget et al. further noted that these post central tumor patients without prompting spontaneously voiced terms such as "will," desire," and "wanting to" during electrical brain stimulation of their inferior parietal lobe regions. As the researchers wrote, these spontaneous, comments offered in real-time by the patients during parietal stimulation conveyed "the voluntary character of the movement intention and its attribution to an internal source, that is, located within the self." (p. 812).

Last, for the pre-central tumor patients, electrical stimulation of the premotor cortex produced very different responses from those recorded in the post-central patients. Here premotor stimulation of Brodmann area 6 triggered movements of various limbs, joints, and the mouth, but without the patients being aware of their actions patients and without the patients showing any evidence of conscious intention. "Patients never expressed the desire to move and never became aware that they produced a motor response" (p. 812). Moreover, varying the intensities of premotor area stimulation did not alter the ability of the patients to detect their electrically-evoked movements. This stood in stark contrast to the effect higher intensities of stimulation of the parietal lobe actually producing the illusion of movement in the patients with post-central tumors. Desmurget et al. thus concluded that inferior parietal lobe sites represent key areas of a widely-distributed neural network that is involved in generating movement intention.

The results of Desmurget et al. extended the earlier work of Fried et al. (1991), which focused on the pre-supplementary motor area, the site of the readiness potential investigated by Libet. Fried and colleagues showed an interesting effect on both subjective experience and actual movement when varying electrical stimulation to this area of the frontal lobe. For low current stimulation, subjects experienced an urge to make a particular movement, whereas for higher intensities subjects produced an actual movement. In an attempt to integrate the findings of Desmurget et al. and Fried et al., Haggard (2009) proposed that conscious intention may entail two distinct processes. One element reflects pre-supplementary motor engagement when subjects focused their attention on initiating a movement, that is, in preparing for the implementation of internally-generated "intentional" action which is then carried out by the primary motor cortex. By contrast, the premotor cortex serves as a staging area for commands for voluntary actions that are being prepared in response to external demands. These premotor signals are then routed to the primary motor cortex for execution. The other aspect of conscious intention, Haggard argued, originates in parietal sites that house a central comparator mechanism that carries out computations to evaluate the fit between predicted and actual sensory feedback. Healthy function of this central comparator mechanism is vital for the establishment of a sense of agency or, as Haggard noted, "…authorship over one's own voluntary movements." (p. 732).

These data are important because they show that the experience of free will reflects the workings of widely-distributed brain areas encompassing a frontal-parietal network whose coordinated activity provides a sense of agency. This network consists of the pre-supplementary area for preparing for the implementation of internally-generated intentional action and the parietal-lobe based central comparator to evaluate the fit between predicted and actual sensory feedback. This represents a neuropsychological model. It is similar to so many others neuropsychological models in which thoughts, feelings and intentions and their corresponding brain processes are given equal weight, and neither is assign primacy and neither is assigned causation. It provides a non-reductionistic and non-deterministic yet materialistic account of free will that is grounded in stochastic, probabilistic relationships between what may be considered, the language of mental life and the language of neural processes. And, as the eminent neuroscientist S.R. Rose noted, just as English can be translated into French, so too can the language of mind be translated into the language of brains, and vice versa, each important for enhancing the understanding of the mechanisms and functions of the other. Such a non-reductionistic, materialistic account may very well be very important for developing bridges between neuroscience of conscious action and criminal responsibility.

7. Insanity jurisprudence

Aristotle arguably planted the seeds for contemporary notions of criminal insanity in his writings on justice and the importance of distinguishing "mad" versus "bad" defendants. Neuroscientists have recently heeded Aristotle's call but have often found themselves ensnared by reductionistic traps of conflating brains and people. People, not brains, have free will. When neuroscientists ascribe an attribute of a person, such as free will, to one of its parts, namely the brain, they commit what philosophers describe as the mereological fallacy (Bennett & Hacker, 2003; Pardo & Patterson, 2010; Rose, 2011). As Pardo and Patterson (2010) argued, neuroscience jurisprudence errs in this "reduction of a psychological attribute to a cortical attribute… a fallacious move from whole to part…" (p. 1226). As these authors argue, "Behavior is something only a human being (or other animal) can engage in. Brain functions and activities are not behaviors (and persons are not their brains)." (p. 1226).

These criticisms are especially important and serve as key reminders of the different goals of neuroscience and law, with the former addressing the question of "is" as in what the brain is made of (fact) and how and why it works (theory), and the latter questions of "ought" as in what ought to be a just standard of personal responsibility, culpability, or punishment (see also Eastman & Campbell, 2006). By the same token, however, as will be discussed below, a careful analysis of the legal statute of criminal responsibility certainly leaves open if not demands the latest advances in neuroscientific knowledge. Moreover, as will also be discussed below, a non-reductionistic yet materialistic neuropsychological model of willed action and agency, as presented above, may help to understand better the relation of brain and criminal responsibility.

8. Brain and criminal responsibility practices

For the legal system, free will is the essential moral cornerstone for establishing mens reas (guilty mind): a defendant is held responsible only if the criminal act occurred freely and with the actor's understanding of its wrongfulness (see Melton, Petrilia, Poythress, & Slobogin, 2007). To determine criminal responsibility requires an investigation and formal mental health assessment of the "mental state at the time of the offense" which entails a reconstruction of a defendant's thoughts,

feelings, perceptions and actions before and during the alleged crime. There are variations of the insanity test, but the most widespread rule, drafted by the American Law Institute (ALI) reads,

> A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality [wrongfulness] of his conduct or to conform his conduct to the requirements of the law.

The ALI standard identifies two prongs, cognitive and volitional, with the former referred to an appreciation of wrongfulness whereas the latter to control of behavior. Impairment in either of these two prongs must be deemed as substantial, that is, a direct result of mental disease or defect. More recently, as of 1995, about 30 states had dropped the volitional prong following the recommendation made by both the American Bar Association (ABA) and the American Psychiatric Association in the aftermath of the acquittal by reason of insanity of John Hinckley in the assassination attempt of President Ronald Reagan (Melton et al., 2007). The ABA contended that conclusions about volition could not be reliably drawn, because the "causes" of a person's behavior cannot be established with any degree of precision. By contrast, greater confidence can be assigned to conclusions about the cognitive prong due to the increased precision and reliability of assessing a person's awareness, perceptions and understanding of an event.

Regardless of which version of insanity test is employed, all adopt a Cartesian model that distinguishes between "the person" and "the disease". This dualism, while standing in contrast to the prevailing monism of neuroscience, is nonetheless in keeping with a medical science model of disease. The emphasis on mental disease in the legal theory of insanity has been often overlooked in scholarship advocating for a neuroscience-informed jurisprudence. In the eyes of the law, a positive neurological finding generated from brain imaging would not meet the legal criteria of mental disease/defect, the essential condition in criminal insanity. For the legal system, actions do indeed speak louder than brain images (Mandavilli, 2006). In fact, more probative to the question of criminal responsibility are symptoms that (a) occur within the context of mental disease, and (b) result in the lack of a substantial capacity either to appreciate the wrongfulness of the alleged behavior or to conform conduct to the requirements of the law. Moreover, with most states eliminating the volitional prong, impulsivity as a defining feature of many brain injuries, most notably frontal lobe disorder would very likely fall below threshold for consideration of insanity.

A large focus of insanity jurisprudence has thus been on the appreciation prong. Here a key factor for the determination of insanity is the extent to which personal agency is compromised by mental disease/defect. Now the question is framed such that neuroscientific research may make a highly relevant contribution to the determination of criminal responsibility. That is, it may offer specialized knowledge to aid the court in moving forward toward justice. However, it does so within the boundaries of the legal system, without threatening legal doctrine, and without reducing fundamental human choices to deterministic brain function. Consider as an example the neuropsychological model of willed action and agency, discussed earlier, which demarcated the joint contributions of frontal motor areas and parietal monitoring areas as key to generating a sense of personal control over one's own actions (Desmurget et al., 2009). Among the advantages of this model is that it demarcates a widely-distributed yet discrete frontal-parietal network whose coordinated activity is hypothesized to enable a sense of agency. If this is correct, then disturbances in a sense of personal agency may be hypothesized or expected to reflect disruptions in the workings of this neural network. In future neuroscience jurisprudence, a claim of insanity based on a person's lack of the substantial capacity to appreciate the wrongfulness of alleged criminal activity may therefore be expected to be characterized by a certain kind of neuropsychological signature that can be assessed by brain imaging techniques. While such neuroscientific import would not be dispositive, it may very well be probative.

9. Mental disease/defect and agency

All legal tests of insanity require at the minimum, or as a threshold, a clinical finding that a defendant suffered from a mental disease or defect at the time of the alleged crime. Legal definitions of mental illness or disease vary across jurisdictions, but all share a restrictive and narrow scope, with all emphasizing that the disorder be of *substantial* proportions and that it leads to *gross* impairment in judgment, perception, reality testing, and everyday functioning [Emphasis added]. The rationale is to target only those rare defendants who genuinely lack agency, who are so grossly impaired as to be robbed of the requisite modicum of rationality to understand and obey the commands of the law (Melton et al., 2007). In practice, what this generally means is that the courts equate mental disease with psychosis, and mental defect with mental retardation (now referred to as intellectual disability), and excluded from this threshold are alcohol- or drug-induced "insanity" and personality disorders, most notably those defined by antisocial behavior (Melton et al., 2007). If this strict threshold is met, then the next required step is that the mental disease or defect "causes" a dysfunction that in turn compromises appreciation of wrongfulness, or control of the criminal acts in those jurisdiction that still allow for the volitional prong.

No particular clinical diagnosis can be equated with insanity or its threshold. In fact, the main focus of an insanity inquiry is not clinical diagnosis per se but rather symptoms, provided that they occur within the context of mental disease/defect meeting criteria of the particular jurisdiction. For example, the most serious of psychotic disorders, schizophrenia, would not necessarily meet legal criteria of mental disease. However, often prominent among the heterogeneous symptoms of schizophrenia are altered experiences of consciousness, specifically in disease-related deficits in self awareness and understanding of willed action and agency (e.g. Blakemore, Wolpert, & Frith, 2002; Brune et al., 2008; Fisher, McCoy, Poole, & Vinogradov, 2008; Franck et al., 2001; Holt et al., 2011; Parnas, Handest, Jansson, & Saebye, 2005). These so-called first-rank symptoms of schizophrenia are distinguished by psychotic experiences of thought insertion, thought control, and passivity phenomena, which are presumed to reflect faulty attributions of thoughts and actions to external sources.

Criminal responsibility presupposes at the minimum an awareness of one's actions, even if lacking an understanding and appreciation of the true source of actions and deeds as originating from within the self (see e.g., Morse, 2013). First-rank symptoms of schizophrenia can compromise this fundamental sense of self-awareness and agency, and patients with these symptoms can often experience their actions, speech, thoughts, or emotions as generated for them by some external agent rather than their own free will (Blakemore et al., 2002). As the Desmurget et al. study of electrical brain stimulation suggested, this healthy and fundamental sense of agency may originate from specific neural interactions coordinated across frontal motor areas and parietal monitoring sites. Disease-related disruptions in this frontal motor and parietal monitoring network have been linked to passivity experience of first-rank symptoms in schizophrenia (Spence, 1996). Moreover, recent experimental studies have pointed to abnormalities in the awareness of action in patients with schizophrenia (Synofzik, Thier, Leube, Schlotterbeck, & Lindner, 2010; Voss et al., 2010). In addition, Voss et al. (2010) reported that deficits in predicting the relationship between an action and its effect correlated with severity of first-rank symptoms, such as delusions and hallucinations.

To summarize, the law views criminal responsibility through the lens of a medical model. Advances in neuroscience may shed light on the issue of criminal responsibility by demonstrating (a) how the experience of intentionality and agency is generated by specific interactions of a discrete network of brain regions of frontal motor and parietal monitoring sites, (b) how mental disease/defect may compromise this network, and (c) how such pathologies may lead to disturbances in the sense of agency that often are central to the phenomenological

experience of schizophrenia. On the other hand, neuroscience will falter if it fails to appreciate (a) the distinct boundaries and goals separating law and scientific inquiry, (b) the inherent limitations of deterministic brain models in accounting for the probabilistic and stochastic essence of human actions, and (c) the perils of overstated and exaggerated claims, which Morse (2006) has dubbed as the "Brain Overclaim Syndrome".

10. Cultural evolution and criminal responsibility

The human brain is tightly constrained by a general architecture that is inscribed genetically and sculpted by millions of years of evolutionary history. Its anatomical structure is universal and its functional capacities are determined by selective pressures of species survival and reproduction. The brain is neither pre-wired at birth nor is it a blank slate with limitless potential in meeting the external demands of the world. Rather the brain, though constricted in scope by biological parameters, is endowed with a modicum of plasticity so that the details of its circuitry and connectivity can be modified in response to culture and environment. Such neural plasticity sets the stage for cultural learning from which the uniquely human ability for pedagogy emerges. As Dehaene (2009) stressed, humans are the only primates to employ pedagogy to actively transmit culture, that is, the shared or communal mental representations, practices, images, and narratives that define a group of human beings. And only humans have the capacity to attend to the knowledge and mental states of others that allows for teaching and propagating cultural representations from brain to brain or from one person to the next.

The genesis of legal systems can be similarly framed, that is emerging through intricate processes of co-evolution of the human brain with cultural minds. Legal systems can thus share universal principles, yet are quite culturally variable in practice. Their adaptive function is to serve as the cultural conscience of a society. Free will represents a core element or building block for the creation of any legal system. From an evolutionary perspective, free will is naturally selected for both its utility as well as for its neural resonance — that is, the ease with which the experience and belief of free will can be either incorporated by preexisting brain circuitry or accommodated by synaptic plasticity or new cortical connections. This "neural niche," to borrow from Dehaene provides free will with the stability for it to be naturally intuited and absorbed rapidly, as well as intentionally transmitted through instruction, teaching, and codified laws. The key ingredients are a brain host that is endowed with properties to both support and constrain this particular form of cultural learning. When these conditions naturally evolve, free will proliferates as a defining feature of a human group, and becomes endemic to systems of social justice.

Criminal responsibility with its presumption of free will serves both social and truth-seeking demands of justice. From a social perspective, it is indeed difficult to conceive of the emergence of partnerships and groups without members sharing in the belief, experience, or intuition that their actions are caused by free will. Likewise, from sociobiology comes the idea that the origins and mechanisms of cooperation trump those of competition in the evolution of social behavior (Nowak & Highfield, 2011). Here the Darwinian mechanism of natural selection expands, choosing social adaptations for an ecological and geographical niche that favor not individual but group fitness (Wilson, 2007). A similar point, offered by De Dreu et al. (2010), proposed that the human brain evolved to sustain motivated cognition and behavior critical to the survival of one's group. The salient point for this paper I would argue is that a defining feature of this sustained cognition and behavior is the belief in free will.

In summary, systems of criminal justice help human societies to maintain important forms of cooperation. The doctrine of criminal responsibility fits with a brain-culture co-evolution model that posits free will as a core element of human cooperation and justice. And such a framework may help to constrain and focus neuroscience on elucidating the clinical underpinnings of criminal responsibility; to wit, how advances in neuroscience might be applied in the service of better scientific understanding and diagnosing of the brain bases for the insanity threshold condition of mental disease/defect. However, for the moral questions of fairness and justice, and for the ultimate legal question of criminal responsibility, these matters are best left to the collective wisdom of the court.

References

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell Press.

Blakemore, S. -J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*, 237–242.

Brown, J. W. (1989). The nature of voluntary action. *Brain and Cognition*, *10*, 105–120.

Brune, M., Lissek, S., Fuchs, N., Witthaus, H., Peter, S., Nicolas, V., ... Tegenthoff, M. (2008). An fMRI theory of mind study in schizophrenic patients with "passivity" symptoms. *Neuropsychologia*, *46*, 1992–2001.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., ... Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, *313*, 1626–1628.

Churchland, P. S. (1990). *Neurophilosophy: Toward a unified science of the mind/brain*. Cambridge, MA: MIT Press.

Crick, F. (1994). *The astonishing hypothesis: The scientific search for the soul*. New York, NY: Touchstone.

De Dreu, C. K. W., Greer, L. L., Handgraff, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., ... Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*, 1408–1411.

Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. New York, NY: Penguin Books.

Desmurget, M., Reilly, K. T., Richard, N., Szathmai, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, *324*, 811–813.

Donald, M. (2010). Consciousness and the freedom to act. In R. F. Baumeister, A. R. Mele, & K. D. Vohs (Eds.), *Free will and consciousness. How might they work?* (pp. 8–23). Oxford: Oxford University Press.

Eastman, N., & Campbell, C. (2006). Neuroscience and legal determination of criminal responsibility. *Nature Reviews Neuroscience*, *7*, 311–318.

Fisher, M., McCoy, K., Poole, J. H., & Vinogradov, S. (2008). Self and other in schizophrenia: A cognitive neuroscience perspective. *American Journal of Psychiatry*, *165*, 1465–1472.

Franck, N., Farmer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, *158*, 454–459.

Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organization of human supplementary motor cortex. *Journal of Neuroscience*, *11*, 3656–3666.

Gallagher, S. (2006). Where's the action? Epiphenomenalism and the problem of free will. In W. Banks, S. Pockett, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation in the nature of volition* (pp. 109–124). Cambridge, MA: MIT Press.

Gazzaniga, M. S. (2005). *The ethical brain: The science of our moral dilemmas*. The Dana Press.

Gazzaniga, M. S. (2011). Neuroscience in the courtroom. *Scientific American*, *304*, 54–59.

Gingerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20–29.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B*, *359*, 1775–1785.

Greene, J. D., & Cohen, J. D. (2006). For the law, neuroscience changes nothing and everything. In S. Zeki, & O. Goodenough (Eds.), *Law and the brain* (pp. 207–226). Oxford: Oxford University Press.

Haggard, P. (2009). The sources of human volition. *Science*, *324*, 731–733.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Holt, J. D., Cassidy, B. S., Andrews-Hanna, J. R., Lee, S. M., Coombs, G., Goff, D. C., ... Moran, J. M. (2011). An anterior-to-posterior shift in midline cortical activity in schizophrenia during self reflection. *Biological Psychiatry*, *69*, 415–423.

Horst, S. (2011). *Laws, mind, and free will*. Cambridge, MA: MIT Press.

Lane, E. (2006). Neuroscience in the courts — A revolution in justice? *Science*, *313*, 458–461.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*, 311–328.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, *8*, 529–566.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential). The unconscious imitation of a freely voluntary act. *Brain*, *106*, 623–642.

Mandavilli, A. (2006). Actions speak louder than images. *Nature*, *444*, 664–665.

Melton, G. B., Petrilia, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts. A handbook for mental health professionals and lawyers* (3rd ed.). New York, NY: The Guilford Press.

Morse, S. J. (2006). Brain overclaim syndrome and criminal responsibility. A diagnostic note. *Ohio State Journal of Criminal Law*, *3*(39), 397–412.

Morse, S. J. (2013). Common criminal law compatibilism. In N. A. Vincent (Ed.), *Neuroscience and legal responsibility* (pp. 27–52). Oxford: Oxford University Press.

Nowak, M. A., & Highfield, R. (2011). *Super cooperators: Altruism, evolution, and why we need each other to succeed.* New York, NY: Free Press.
O'Hara, E. A. (2004). How neuroscience might advance the law. *Philosophical Transactions of the Royal Society of London B*, *359*, 1677–1684.
Pardo, M. S., & Patterson, D. (2010). Philosophical foundations of law and neuroscience. *University of Illinois Law Review*, *4*, 1211–1250.
Parnas, J., Handest, P., Jansson, L., & Saebye, D. (2005). Anomalous subjective experience among first-admitted schizophrenia spectrum patients: Empirical investigation. *Psychopathology*, *38*, 259–267.
Piatelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds.* New York, NY: John Wiley & Sons, Inc.
Rigoni, D., Kuhn, S., Sartori, G., & Brass, M. (2011). Inducing disbelief in free will alters brain correlates of preconscious motor preparation: The brain minds whether we believe in free will or not. *Psychological Science*, *22*, 613–618.
Rose, S. (2011). Self comes to mind: Constructing the conscious brain by Antonio Damasio — Review. *The Guardian* (February 12, 2011).
Roskies, A. L. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences*, *10*, 419–423.
Roskies, A. L. (2010). Freedom, neural mechanism, and consciousness. In R. F. Baumeister, A. R. Mele, & K. D. Vohs (Eds.), *Free will and consciousness. How might they work?* (pp. 153–171). Oxford: Oxford University Press.
Searle, J. R. (2010). Consciousness and the problem of free will. In R. F. Baumeister, A. R. Mele, & K. D. Vohs (Eds.), *Free will and consciousness. How might they work?* (pp. 121–134). Oxford: Oxford University Press.
Spence, S. A. (1996). Free will in the light of psychiatry. *Philosophy, Psychiatry, and Psychology*, *3*, 75–90.
Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain*, *133*, 262–271.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
Vohs, K. D. (2010). Free will is costly: Action control, making choices, mental time travel, and impression management use precious volitional resources. In R. F. Baumeister, A. R. Mele, & K. D. Vohs (Eds.), *Free will and consciousness. How might they work?* (pp. 66–81). Oxford: Oxford University Press.
Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: A specific deficit in predicting action consequences. *Brain*, *133*, 3104–3112.
Walter, W. G., Cooper, R., Aldrige, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, *203*, 380–384.
Wegner, D. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.
Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of free will. *American Psychologist*, *54*, 480–491.
Wilson, T. D. (2002). *Strangers to ourselves. Discovering the adaptive unconscious.* Cambridge, MA: The Belknap Press of Harvard University Press.
Wilson, D. S. (2007). *Evolution for everyone: How Darwin's theory can change the way we think about our lives.* New York, NY: Delacorte Press.
Zeki, S., & Goodenough, O. (2004). Law and the brain: introduction. *Philos. Trans. R. Soc., B Biol. Sci.*, *359*(1451), 1661–1665. https://doi.org/10.1098/rstb.2004.1553.
Zeki, S., & Goodenough, O. (2006). *Law and the brain.* Oxford: Oxford University Press.