

REVIEW ARTICLE

Open Access

The charm of structural neuroimaging in insanity evaluations: guidelines to avoid misinterpretation of the findings

C. Scarpazza^{1,2}, S. Ferracuti³, A. Miolla¹ and G. Sartori¹

Abstract

Despite the popularity of structural neuroimaging techniques in twenty-first-century research, its results have had limited translational impact in real-world settings, where inferences need to be made at the individual level. Structural neuroimaging methods are now introduced frequently to aid in assessing defendants for insanity in criminal forensic evaluations, with the aim of providing “convergence” of evidence on the *mens rea* of the defendant. This approach may provide pivotal support for judges’ decisions. Although neuroimaging aims to reduce uncertainty and controversies in legal settings and to increase the objectivity of criminal rulings, the application of structural neuroimaging in forensic settings is hampered by cognitive biases in the evaluation of evidence that lead to misinterpretation of the imaging results. It is thus increasingly important to have clear guidelines on the correct ways to apply and interpret neuroimaging evidence. In the current paper, we review the literature concerning structural neuroimaging in court settings with the aim of identifying rules for its correct application and interpretation. These rules, which aim to decrease the risk of biases, focus on the importance of (i) descriptive diagnoses, (ii) anatomo-clinical correlation, (iii) brain plasticity and (iv) avoiding logical fallacies, such as reverse inference. In addition, through the analysis of real forensic cases, we describe errors frequently observed due to incorrect interpretations of imaging. Clear guidelines for both the correct circumstances for introducing neuroimaging and its eventual interpretation are defined

Introduction

Despite the popularity of structural neuroimaging techniques in twenty-first-century research, its results have had limited translational impact in real-world settings, where inferences need to be made at the individual level. Recently, however, a growing number of research groups have attempted to overcome this issue by studying the feasibility of single-case analysis^{1–6}.

Structural neuroimaging (sNI) evidence is increasingly being used in criminal trials^{7–9}. The literature on sNI and law tends to consider neuroscientific knowledge extremely relevant for assisting in psychiatric assessments of criminal responsibility^{10–12}. In the penal field, holding a defendant criminally responsible requires proof that he/she committed the act (*actus reus*) and that the act was intentional, i.e., the defendant committed the act of his/her own free will (*mens rea* or guilty mind)¹³. Crucially, the *mens rea* can be abolished (or weakened) due to the impact of a medical condition on brain functioning¹⁰, usually a neurologic or psychiatric condition (or, more rarely, organic conditions, such as hypoglycemia¹⁴). In these cases, the defendant could be considered mentally insane at the moment of the crime and, consequently, not responsible or having diminished responsibility for his/

Correspondence: C. Scarpazza (Cristina.scarpazza@gmail.com) or (cristina.scarpazza@unipd.it)

¹Department of General Psychology, University of Padua, Via Venezia 8, 35131 Padova, Italy

²Department of Psychosis Studies, Institute of Psychiatry, Psychosis and Neuroscience, King’s College London, De Crespigny Park, London SE5 8AF, UK Full list of author information is available at the end of the article.

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

her actions. Thus, the determination of criminal responsibility rests on the evaluation of the mental state of the defendant at the time of the crime¹⁵.

Critically, a wide body of recent literature has revealed that forensic insanity assessment is particularly prone to different biases^{16,17}. Most relevant is the so-called “cognitive bias”, arising from the subjective humanness that influences every step of the forensic psychological assessment^{18,19}, from data collection to judicial decisions. These biases arise as forensic experts are human beings with personal knowledge, beliefs, expectations, cognitive schemas, etc. All of these factors influence the way forensic elements are addressed and unconsciously impact the collection, analysis, interpretation and reporting of data^{18,20,21}. The hierarchy of expert performance²² (HEP) has recently been proposed to better understand expert performance in forensic science and the biases that might arise at each step of forensic work. The HEP has also been applied to forensic psychology and psychiatry²³, and some bias-mitigating strategies have been proposed²⁰. The HEP makes a clear distinction between observations and conclusions, where observations pertain to data observation, collection and analysis, while conclusions depend on the assessment and interpretation of observations^{22,23}.

Forensic psychologists and psychiatrists are increasingly aware of biases that may influence expert performances^{17,21}. Three critical issues hampering the credibility of forensic experts are as follows: the tendency of experts to be biased by previous expectations²⁴, the low inter-rater agreement in conclusions on insanity based on clinical assessments^{25–28} and the variability in the rates of evaluation outcomes as a function of the evaluator^{17,28,29}. For these reasons, brain images are now increasingly introduced to assist in the assessment of insanity in criminal cases, with the aim of providing a “convergence” of evidence (including anamnestic, clinical and neuroscientific information). This is of critical relevance, as the *mens rea* of the defendant is pivotal to the judge’s decision^{30,31}. The idea is that sNI would provide additional biological data that, when considered in conjunction with classical neurologic, psychiatric and neuropsychologic assessments, would inform the court about the defendant’s responsibility¹¹. The literature suggests that sNI evidence should not be used alone and cannot be used solely to explain the cause of a violent crime (e.g., “she killed her mother because she has a frontal lobe lesion”). Rather, sNI findings (i.e., the frontal lobe lesion) should be considered a “hard” correlate of mental disease, symptoms of which (e.g., disinhibition) are causally linked to the crime (e.g., the homicide)¹¹. Thus, the criteria for responsibility are currently behavioural and should remain that way^{11,12}.

Although the use of sNI in court seems to be a promising advance in reducing the uncertainty of forensic

experts’ evaluations³², their use is not free from potential risks and biases^{11,12,33,34}. One such bias is that neuroimaging results are considered to be objective, which is not always true. In some cases, MRI data need to be analysed with sophisticated techniques^{35,36}, and the results can be influenced by different technical choices. This bias concerns the HEP observation level and will not be explored in the current paper. Another worrying concern, at the HEP conclusion level, is the potential problem of the “inferential distance”^{37,38}, i.e., the vivid and appealing visual nature of neuroimaging that could make jurors think that the images corresponded to objective findings. Interestingly, McCabe and Castel (2008)³⁹ examined whether brain images affect people’s judgements of scientific credibility. They found that when neuroimages accompanied scientific summaries, the summaries were rated as more scientifically credible than summaries presented alone or paired with graphs. Similar results were obtained by Weisberg et al. (2008)⁴⁰, who demonstrated that jurors become more likely to be convinced by logically irrelevant explanations for behaviours if these explanations are simply expressed in terms of neuroscientific evidence. Although these results do not appear to be replicated in a real forensic setting⁴¹, they provocatively underline the seductive charm of neuroimaging. This is a clear example of the biasing influence of human involvement on judicial decisions¹⁸: if sNI evidence is not properly weighed and integrated with other evidence, it is not helpful to the forensic process. This can lead to a potential improper use of sNI in court⁴², where the risk is that consultants are more interested in generating fascinating images of the brain than they are in elucidating their importance for explaining the behaviours that are legally relevant³³.

For this reason, it is important that neuroscientists understand how to correctly interpret sNI evidence in the process of evaluating insanity^{33,34}. Furthermore, it is important for judges and jurors to have sufficient knowledge of the ways in which sNI evidence can be relevant to legal questions and to recognize when it is not. Thus, the aim of the current paper is to present practical guidelines to be used in the delicate process of sNI result interpretation by the following (HEP level of conclusions): (i) forensic neuroscientists, when using neuroimaging as scientific evidence in support of mental incompetence; (ii) judges and jurors, during the sophisticated decision-making process of determining the role of neuroimaging data in support of the criminal liability of defendants. Using the HEP terminology^{22,23}, the guidelines aim to enhance intra-rater and inter-rater reliability and minimize bias in conclusions.

Before discussing guidelines, we will introduce relevant concepts via two real-life forensic cases in which neuroscientific evidence was used to support a clinical

diagnosis. They were selected mainly due to their similarities: both homicide cases, both likely due to disinhibition, and both with defendants manifesting frontal lobe dysfunction. However, there is one important difference between these cases. In the first, an objective brain lesion was evident, while in the second, there was a subtle alteration in grey matter volume. Thus, in the first case, the forensic discussion focused on the causal link between the brain lesion and the homicide (HEP level of conclusions), while in the second case, the discussion focused both on the investigation of brain abnormalities (HEP level of observations, which will not be discussed in detail) and on the anatomo-clinical correlation between the brain abnormalities and behavioural dysfunction (HEP level of conclusions).

As additional premises, we would like to underline that (i) this paper is focused on the use of structural images only. The introduction of functional images is further complicated by high between-subject variability in the blood oxygenation level-dependent (BOLD) signal, the ecological validity of task performance⁴³ and the absence of false-positive distribution maps that can guide result interpretation (this map is available for sNI^{4,5}). Thus, the potential application of functional imaging is a topic that falls outside the aims of the current paper. (ii) In this paper, we have purposely decided to present neuroimaging findings in an accessible and charming way in an attempt to give the readers an idea of their charming impact.

Examples of Real Forensic Application

Case 1

The facts

The defendant was a 40-plus-year-old man charged with murder. In 2014, the man saw a jogger on the street and suddenly stopped and assaulted the woman, who reacted. After having beaten the woman severely, he threw the body off a cliff. The next day, he went back to the scene of the crime and masturbated on the corpse. After ~3 months, he was interrogated by the police and immediately confessed the murder.

The descriptive diagnosis

The defendant was a 40-plus-year-old man who grew up in a rural mountain area. His parents divorced when he was nine. He had completed 5 years of education. He was married and had a son. A few years into the marriage, he suffered from severe head trauma due to a car accident and was hospitalized for 3 months. His marriage lasted <2 years after his accident, and no relationship with the son was maintained. He was unable to keep a job for more than a few months since the accident. Moreover, he developed a serial pattern of criminal sexual behaviour. He was previously jailed for 1 year due to performing

frottage and making vituperative comments to women on the street.

At the forensic psychiatric examination, no psychotic symptoms were evident, but the defendant did appear fatuous and avoidant. He seemed to not realize the severity or legal consequences of his behaviour. The neuropsychological examination revealed a borderline IQ, impaired verbal comprehension, and below-average working memory function. The defendant presented with severe frontal executive functioning deficits, in particular in the ability to inhibit automatic answers (as measured using the Stroop test⁴⁴) and behaviours (as measured using the frontal assessment battery⁴⁵). During the neuropsychological evaluation, his behaviour was oppositional; he was easily frustrated with negative feedback and had a tendency to become irritated and verbally aggressive towards the examiner.

The brain diagnosis

The CT scan (MRI could not be acquired due to the defendant's pacemaker) showed a brain lesion, which was likely a result of the head trauma that had occurred 10 years prior. The lesion was detected bilaterally in the mesial prefrontal cortex, measuring 12.29 cm³ (Fig. 1).

Interpretation of the results

According to the judge's expert, the neuroimaging data in this criminal case were fundamental in explaining the constellation of symptoms that characterized the defendant's history, as well as the crimino-dynamics that led to the homicide. From the defendant's anamnesis, it is evident that he had undergone a "behavioural fracture" from being a man with a normal life, marriage, and job to exhibiting sexually depraved/disinhibited behaviour and an inability to maintain normal relationships or jobs. Moreover, his sexual behaviour was carried out in a

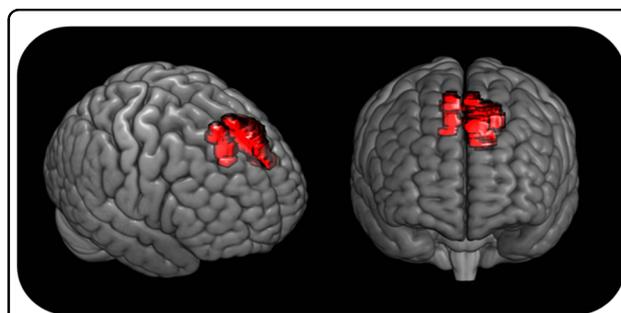


Fig. 1 3D reconstruction of the defendant's exact brain lesion.

Based on the CT scan, the brain lesion was manually traced by one of the authors (CS) on the MRICro template. The traced lesion was then overlaid to the 3D render provided by MRICroGL software (<http://www.cabiatl.com/mricrogl/>)

disorganized and risky manner. The man used to perform frottage on women in broad daylight, and in the case of the murder, the aggression occurred near a main road with a high probability of being seen.

Critically, the defendant manifested additional behavioural symptoms that were a strong indication of an affected ability *to understand* and *to will*. According to the Italian penal code, these two capabilities are the legal basis of the insanity plea. Regarding the defendant's ability to understand, he seemed not to realize the legal consequences of his actions; he did not understand the risk of being caught (supported by his return to the scene of the crime), and he manifested a deficit in moral reasoning. Regarding the defendant's ability to will, his personal history clearly revealed previous aggressions towards women, previous outbreaks of violence, and repeated inability to inhibit sexual urges. All these symptoms strongly indicated pathology of the prefrontal cortex. Even more interestingly, he manifested symptoms that were legally irrelevant *per se* but strongly support frontal lobe dysfunction through the "convergence of evidence" principle. For instance, he was unable to inhibit non-sexual impulses, as emerged from the neuropsychological examination and from his verbally aggressive behaviour during said evaluations.

The presence of the brain lesion is also important from the crimino-dynamic prospective: after the traumatic brain injury, the defendant manifested sexually deviant behaviours mainly characterized by frotteurism. Sexual alterations after traumatic brain injury are well known in the literature^{46,47}. Thus, there is a strong cause for causality between the brain lesion and the alteration in sexual behaviour. However, this is not sufficient to explain the murder. The defendant's violent behaviour may be explained by his easy frustration in response to negative feedback (the woman's refusal). This frustration was followed by aggressive and disinhibited behaviour (the strike), which found its equivalent in verbal aggression towards the examiner during the neuropsychological examination. Again, disinhibition is strongly suggestive of frontal lobe dysfunction^{48–50}. Notably, the defendant's manifest insensitivity to the consequences of behaviour corroborated the idea of frontal lobe deficits⁵¹. Thus, a strong anatomo-clinical correlation between the brain lesions and pathological symptoms of the defendant is evident.

In summary, in this case, the judge's expert concluded that the constellation of symptoms, legally relevant or not, could be explained by the presence and location of the traumatic brain lesion. The criteria for mental insanity are grounded in the defendant's behaviour, and neuroimaging findings provide a useful explanation for them. The judge concluded that the defendant should thus be considered not responsible for his behaviour.

What if I get it wrong?

Gianfranco Stevanin was an Italian serial killer⁴³. He suffocated six prostitutes during extreme sexual activities and then dismembered and buried their bodies. The man had suffered a severe head injury resulting in a bilateral lesion of the frontal lobes (no images are available). From the time of the brain injury, he manifested asphyxiophilic tendencies⁵². Because the frontal lobes are the brain regions responsible for impulse control and inhibition, the man was initially declared not responsible for the murders. According to the defence thesis, as a result of the bilateral frontal lobe lesion, the man was unable to inhibit sexual activity once it began. Corroborating this hypothesis, the neuropsychological evaluation revealed the presence of many symptoms that clearly indicate frontal lobe dysfunction: disadvantageous decision-making, impaired impulse inhibition and easy irritability. In contrast, his behaviour after the murders, e.g., the dismemberment and concealment of the dead bodies, could not be explained by a deficit in impulse inhibition. Critically, it was then discovered that Gianfranco Stevanin used to practise the very same extreme sexual activities with his fiancée, without killing her. These actions provided unequivocal evidence discrediting the claim that the man could not interrupt his sexual impulses. In this case, the presence of the brain injury could not be used to mitigate the man's responsibility for his crimes. Gianfranco Stevanin's case is of outstanding importance to exemplify that, even in the presence of a clear brain lesion, the interpretation of its impact on behaviour is "somewhat subjective" (human element¹⁸) and that clinicians should consider all clinical and behavioural information to correctly interpret the meaning of the imaging findings.

Case 2

The facts

The defendant was a 24-year-old woman, JF, who was charged with murder for smothering her newborn child to death immediately after delivery³¹. She then wrapped the infant's body in a towel and hid it inside a suitcase. The defendant later claimed that the newborn child was 'born dead' due to drug abstinence syndrome (see below).

The descriptive diagnosis

Anamnestic information revealed that the defendant started to heavily smoke cigarettes at the age of eleven. She had a well-documented history of multidrug abuse, as well as alcohol abuse, since the age of thirteen. She had become pregnant during a party and did not interrupt her drug use during pregnancy.

The forensic psychiatric examination revealed that the defendant had a personality profile characterized by antisocial features such as a history of illegal behaviour, sensation seeking, familial conflict, lack of sensitivity,

rejection of conventional standards, poor response to threatening situations and a willingness to forgo careful consideration of alternative solutions to problems. The neuropsychological evaluation revealed impulsivity, i.e., inability to inhibit the prepotent response (as measured by means of the Hayling test⁵³); a deficit in planning (as measured by means of the Tower of London⁵⁴); and deficits in emotional attribution and in identifying violations of social norms (as measured by the emotion attribution task⁵⁵ and social situation task⁵⁶).

The brain diagnosis

The structural MRI of the defendant was compared with the MRI of healthy women using voxel-based morphometry (VBM), a neuroimaging technique that highlights subtle structural anatomical abnormalities^{35,36}. The analysis revealed a reduced grey matter volume in the left prefrontal cortex in JF relative to the control group³¹ (Fig. 2).

Interpretation of the results

According to the defence consultant, the sNI data in this case were fundamental in explaining the constellation of symptoms that characterized the defendant’s history. Indeed, the site of the brain alteration (i.e., the frontal lobe) has a strong anatomo-clinical correlation with the descriptive diagnosis, which can be summarized as frontal syndrome, as the defendant manifested symptoms ascribable to frontal lobe dysfunction, particularly impulsivity⁵⁷, a deficit in planning, a deficit in automatic response inhibition^{58,59} and difficulty in emotional attribution⁶⁰, as well as alteration of the personality characterized by high scores on the borderline⁶¹, depression⁶² and substance abuse⁶³ subscales. Furthermore, according

to the false-positive distribution map³, the decrease in grey matter in the dorsolateral prefrontal cortex (DLPFC) in a single individual compared to the control group has only a 0.5–5% probability of being a false positive.

The presence of the brain alteration in the frontal lobe is also important from the crimino-dynamic prospective: the symptoms identified in JF showed that she had a reduced capacity to control her behaviour. This was extremely relevant for the evaluation of the prosecutor’s hypothesis, according to which JF *impulsively* smothered the newborn to death. Thus, the clinically identified pathological features are causally linked to the crime, providing the basis for an insanity defence³¹. JF’s behavioural reports, accompanied by reduced grey matter in the frontal lobe, provided “hard” biological evidence for the defence’s case.

In summary, the defence consultants concluded that the constellation of behavioural and neuropsychological symptoms manifested by JF were in strong anatomo-clinical correlation with the brain abnormalities and were causally linked to the crime. According to the defence consultants, JF should be considered not responsible for her behaviour because she was unable to inhibit her impulses.

What if I get it wrong?

A woman was charged with murder for running over a man with her car. She underwent psychiatric and psychological evaluations to understand her state of mind at the moment of the homicide. The descriptive diagnosis gave a negative result. The defendant’s brain MRI, analysed using VBM methodology, revealed an increase in grey matter in the DLPFC, particularly in the middle frontal gyrus (Fig. 3, left panel). In the absence of any descriptive diagnosis, this result can be interpreted only according to the literature. Since an increase in DLPFC grey matter has been linked to binge drinking⁶⁴ (Fig. 3, right panel), it would be tempting to infer that the defendant was a binge drinker and that the accident occurred because she was drunk.

This conclusion and reasoning are invalid, since inferring the presence of altered mental states (being a binge

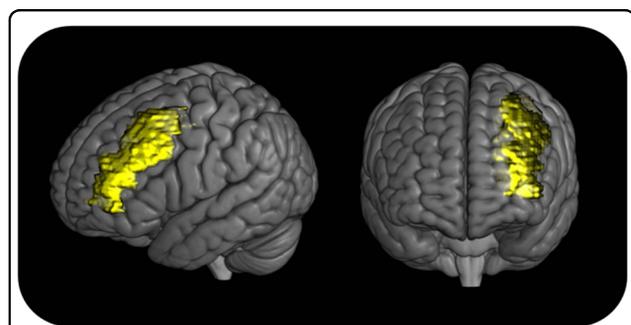


Fig. 2 3D reconstruction of the localization of JF’s brain deficiency in grey matter volume. In the image, the exact extent of JF’s abnormality is not represented, but only the location. In the reference paper³¹, the grey matter loss is reported to be localized in the middle frontal gyrus. For this image, a binary mask on the whole middle frontal gyrus was created using PickAtlas (<http://fmri.wfubmc.edu/software/pickatlas>) and was then overlaid onto the 3D render provided by MRICroGL software (<http://www.cabiatl.com/mricrogl/>)

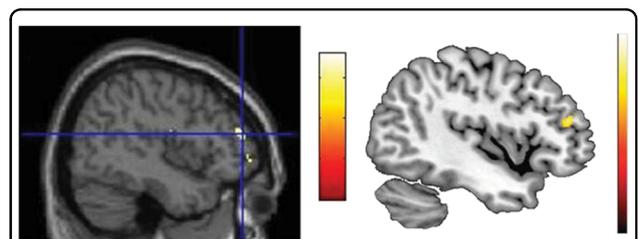


Fig. 3 VBM results. Reprinted from ref. ⁴. Left panel: VBM result of the defendant’s brain. Right panel: VBM results in binge drinking; image adapted from ref. ⁶⁴

drinker) from the presence of a brain alteration (increased grey matter in the DLPFC) is a logical fallacy called reverse inference⁶⁵. This fallacy is particularly relevant in neuroimaging studies, where behaviours are subserved by complex brain networks and each brain region is consequently involved in more than one behaviour⁶⁶. Furthermore, according to the false-positive distribution map⁶, an increase in grey matter in the DLPFC in a single individual compared to the control group has up to a 23% chance of being a false positive. Thus, this result, which has no clinical correlate (i.e., is not in anatomic-clinical correlation with the descriptive diagnosis), is likely to be a false positive result, probably reflecting normal variability in neuroanatomy rather than any brain pathology⁶. Indeed, this case was fictional, and the VBM results reported highlighted the grey matter increase in the brain of one of the authors (CS) compared to a control group.

Discussion

In the current paper, we aim to provide clear guidelines for the correct use of sNI in court. Our opinion is that sNI should support behavioural findings to reduce controversies in court. Both of the described cases clearly demonstrate the importance of using neuroimaging findings strictly in conjunction with behavioural evidence. Based on the danger of incorrectly claiming the presence or absence of a causal relationship between neuroimaging findings and behaviour, we believe that careful behavioural assessment remains more reliable than neuroimaging in ascertaining the relevant mental states that form the basis of legal criteria.

While it is certainly true that neuroimaging is a powerful tool whose potential for misuse can potentially cause deleterious effects in a legal setting⁶⁷, we believe it is a mistake to deny its potential helpful applications a priori^{12,68}. Although neuroscientists often commit logical errors when presenting isolated neuroscientific findings^{33,67}, we believe a similar logical error is committed by ignoring neuroscientific data⁶⁹. It is our opinion that the legal and scientific communities should work together to create guidelines that ensure the correct application of neuroscientific techniques and the appropriate interpretation of neuroscientific findings. In the following paragraphs, we will discuss general rules for the use of neuroimaging. The correct application of these rules is extremely important to allow group-to-individual inference, a problem that neuroscientists recognize and are trying to solve^{2-6,67}. Finally, we will propose guidelines that might be useful to neuroscientists, lawyers, judges and jurors. These rules and guidelines will help to increase within- and between-expert reliability and to reduce within- and between-expert bias in expert testimony conclusions^{22,23}.

On the role of neuroimaging in court

Rule number 1. Neuroimaging results should be coupled with behavioural findings. Neuroimaging results assume a meaning only when coupled with a clear descriptive diagnosis, i.e., with clear symptoms manifested by the defendant. Indeed, without a clear descriptive diagnosis, the neuroimaging findings cannot be interpreted, as described in the paragraph “What if I get it wrong?” of case 2. This rule has been violated in numerous cases reported in the literature. One example is the famous case of Weinstein, where the defence attorney used the neuroimaging findings of a subarachnoid cyst in the frontal lobe to support the insanity of the defendant^{67,70}. Although the presence of arachnoid cysts has been associated with atypical psychosis⁷¹, no one tested for the presence of psychosis in the defendant or his inhibitory abilities, moral reasoning, etc. We recognize that this rule may open the door to biased neuroimaging results because there is a motive to couple them with behavioural findings. Although this is true, this potential bias lies at the observation level of HEP and can be overcome by blinding the experts to the descriptive diagnosis of the defendant^{20,72}. Rule number 1 applies mainly to wider forensic reasoning (HEP conclusions), when the attorneys need to integrate evidence.

Rule number 2. The criminal behaviour cannot be considered a symptom. Anatomic-clinical correlation can be assessed between a brain region and cognitive function but cannot be assessed between a brain region and criminal behaviour, as there is no specific brain region involved in complex behaviours such as criminality or violence. Despite investigation of this topic in recent studies⁷³, evidence suggests that complex behaviours are supported by highly complex brain networks^{73,74}. Complex behaviours can be broken down into contributing cognitive functions, i.e., an irresistible impulse murder may be partially explained by a deficit in impulse control. Thus, the hypothetical brain abnormality must be put in anatomic-clinical correlation with the lack of impulse control, which should be demonstrated clinically (rule number 1) and, if possible, by evidence beyond the criminal act. For example, we recently studied a case of acquired paedophilia^{12,68} and claimed that the defendant was unable to inhibit his impulses. To support this hypothesis, we referred to the man’s recently emerged kleptomania, a fact that is irrelevant to paedophilia but relevant to impulse control. Rule number 2 was violated in the case of Weinstein, as the defence attorney argued the anatomic-clinical correlation between the cyst in the frontal lobe and the violent murder, using the violent murder as a symptom of the cyst^{67,70}. This rule would help to minimize the adversarial allegiance bias^{17,23}, which refers to the tendency of experts to reach conclusions supporting the side that retained them.

Rule number 3. Not every brain abnormality leads to behavioural symptoms. The correlation between brain function and criminal behaviour is far from perfect. Although violent behaviour is more likely in patients with frontal lobe deficits than in patients with (for example) parietal or occipital deficits, not every patient with frontal lobe damage manifests violent behaviour. In addition to the famous case of Phineas Gage, the poet Guillaume Apollinaire should be considered. After being wounded in the right temporal lobe during World War I, Apollinaire became unaffectionate and indifferent to his girlfriend, his personality became defiant and nostalgic, and his work became darker—but he never manifested aggressive behaviour⁷⁵. Indeed, the brain is incredibly plastic, and different clinical conditions demonstrate that even large brain lesions may not lead to important behavioural deficits. An example is Rasmussen encephalopathy, a rare inflammatory brain disorder leading to the death of one hemisphere⁷⁶. To date, the only known effective treatment is a hemispherectomy. The literature reveals that if a hemispherectomy is performed during childhood, patients may not have severe neurologic sequelae⁷⁶. Rule number 3 has been violated in the case of Gianfranco Stevanin, as the defence attorney wrongly ascribed his violent behaviour to the frontal lobe deficit. This rule would help in

minimizing confirmation bias²⁰, which describes the tendency of experts to reach conclusions that confirm what they already believe to be true.

Rule number 4. Do not reason backwards. To infer the presence of an altered mental state from the presence of brain pathology is a *reverse inference*. Reverse inferences are a logical fallacy that should be avoided^{16,65}. The correct neuroscientific reasoning is the following: the defendant behaviourally manifested difficulties in behavioural control; the defendant also has a brain lesion in the frontal lobe; thus, the brain lesion might account for the lack of behavioural control. Assuming the presence of a psychiatric or neurological disorder from brain images is dangerous both from the clinical (i.e., wrong drug treatment) and forensic (i.e., incorrect conclusion on insanity) points of view. Rule number 4 was violated in both Weinstein’s case and the case described in the “What if I get it wrong?” paragraph of case 2. As with rule 1, rule 4 can be applied to both the observation and conclusion levels of HEP. One way to mitigate the bias of backwards reasoning at the HEP level of observation is to avoid providing neuroimaging experts with irrelevant case information, which might cause cognitive contamination²⁰. One way to mitigate the risk of bias at the HEP level of conclusions is to request neuroimaging

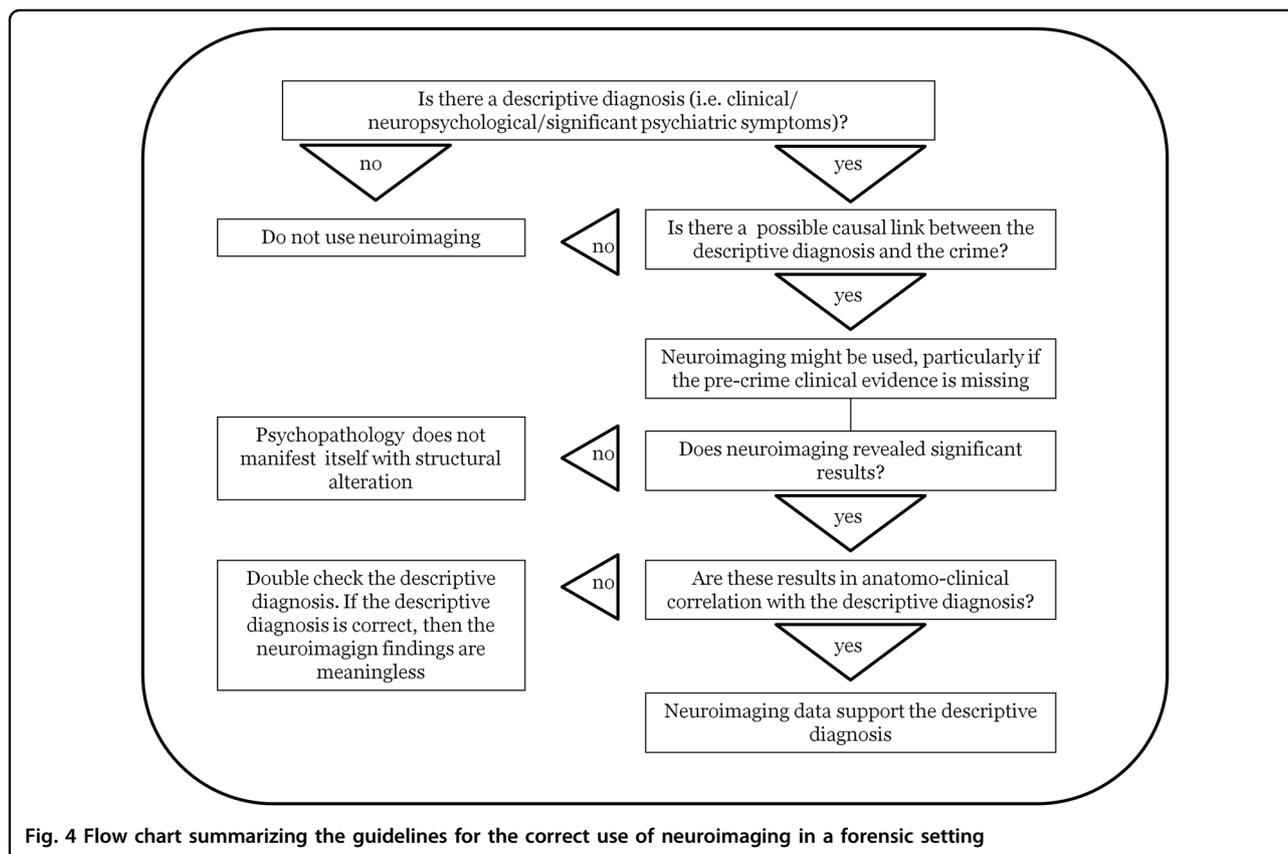


Fig. 4 Flow chart summarizing the guidelines for the correct use of neuroimaging in a forensic setting

investigation only when the defendant presents with clear psychiatric or cognitive symptoms, as explained in the following guidelines.

Practical guidelines to avoid misinterpretation of results

In the current paragraph, we will operationalize the rules in practical guidelines for the correct use of neuroimaging in court. The aim of these guidelines is to mitigate the impact of cognitive biases¹⁸ and the misinterpretation of the neuroimaging results often presented in forensic trials⁶⁷. In particular, the correct application of these guidelines will help to minimize both within- and between-expert biases in conclusions²³, including the adversarial allegiance bias^{17,23}. The guidelines are graphically represented in Fig. 4.

- a. Provide a descriptive diagnosis, i.e., describe the symptoms manifested by the defendant, as behavioural evidence is the gold standard for determining functional impairment. The descriptive diagnosis should be carried out, when possible, with standardized neuropsychological tests including indexes for controlling malingering and exaggeration (e.g., symptom validity testing⁷⁷). If the descriptive diagnosis is negative, as in the fourth case described in the current paper, then any accompanying sNI is useless and should thus be excluded. If the descriptive diagnosis is positive, investigation can proceed to step 2.
- b. Assess the causal link between the symptoms and the crime. The causal link between altered state of mind (i.e., symptoms) and crime⁷⁸ has not been discussed in the current paper, as it was beyond the scope of our topic. However, it is important to remember that in the forensic setting, the symptoms of the defendant are important if they can be causally linked to the actus reus. For instance, impulse disinhibition is of great relevance in the context of the irresistible impulse to murder, while it is not relevant in the context of premeditation. If there is no causal link between the descriptive diagnosis and the crime, then sNI should not be used, as there is no way for eventual results to be interpreted. If there is a causal link between the descriptive diagnosis and the crime, sNI evidence might be useful, particularly if pre-crime clinical evidence is missing.
- c. Clarify whether the neuroimaging techniques highlight significant results. In the case that neuroimaging techniques produce no results, it can be concluded that the psychopathology does not manifest itself with structural alterations. It is important to remember that the presence/absence of psychopathology can be deduced only by behavioural findings (point a); thus, a negative sNI

result does not imply the absence of psychopathology. The identification of neural correlates of psychiatric or personality disorders is highly complicated and may be more evident at the functional level than at the structural level⁷⁹. Of course, if malingering is suspected, negative neuroimaging results could prompt a search for additional behavioural evidence. If sNI results are positive, the investigation can proceed to step d.

- d. Use brain imaging to assess anatomo-clinical correlation. Are the neuroimaging results in anatomo-clinical correlation with the descriptive diagnosis? Although neuroscientists agree that complex behaviours are the result of a dynamic network of interconnected brain regions^{80,81}, it is also true that some cognitive functions rely strongly on a specific brain region: for instance, behavioural control requires the correct functioning of the prefrontal cortex⁸². Thus, a frontal lobe deficit should be expected in a defendant manifesting behavioural impulse disinhibition. If there is no anatomo-clinical correlation, then we suggest that the descriptive diagnosis be checked again. If the descriptive diagnosis is correct, then the sNI results, despite their presence, are not relevant to explain the defendant's behaviour, as in the case of Gianfranco Stevanin. This may be attributable to brain plasticity. Only if there is anatomo-clinical correlation can the sNI data be used to support the descriptive diagnosis.

Limitations

It is important to acknowledge some potential limitations of the current paper. First, we are referring only to structural neuroimaging findings, as we purposely omit discussion of functional neuroimaging for the reasons described in the introduction. Thus, our rules and guidelines cannot be generalized to the use of functional images in court. The use of functional imaging in this setting is, in our opinion, still premature. Second, we neglected to discuss the within- and between-expert biases at the observation level of the HEP model^{22,23}, as there are different ways in which images can be analysed, and the literature on the topic is so wide that it would have been impossible to summarize in the current paper. We are aware that apparent unreliability at the level of conclusions may lie at the level of observation, and this would require a different intervention²³. Future studies are needed to explore this topic.

Conclusions

In the current paper, we used real forensic cases to highlight how neuroimaging results should be interpreted when used as an additional instrument to reduce

cognitive bias at the HEP conclusion level^{22,23} and thus reduce controversies in the assessment of insanity. Through a detailed analysis of two cases and an explanation of errors that one may commit if results are not correctly interpreted, we offered four general rules for the use of neuroimaging in court. Finally, we proposed guidelines for the correct use of neuroimaging in a forensic setting. Our aim was to provide neuroscientists with a practical guide when facing the difficulty of using neuroimaging as scientific evidence in support of a mental incompetence claim.

Although the present study focuses on the use of sNI in a forensic setting, the insights from the current analysis, as well as the rules and guidelines defined, might be of interest for a broad range of neuroscientists.

Indeed:

We are claiming that sNI results not coupled with behavioural findings are meaningless. Thus, this paper may be of interest for disentangling the controversies on the general role of sNI results^{12,67,69}.

The rules and guidelines proposed could be adopted outside forensic contexts and could be applied to individual cases, for instance, in clinical contexts^{2–6}.

The current study provides important, practical and exploitable guidelines that can ultimately be of immediate use to clinicians and neuroscientists who are interested in the real-world application of neuroscientific findings.

Acknowledgements

We are grateful to Kirsten Lee, King's College London, for editing an early version of this manuscript.

Author details

¹Department of General Psychology, University of Padua, Via Venezia 8, 35131 Padova, Italy. ²Department of Psychosis Studies, Institute of Psychiatry, Psychosis and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK. ³Department of Human Neuroscience, Sapienza University of Rome, P.le A Moro 5, 00185 Roma, Italy

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 26 February 2018 Revised: 24 August 2018 Accepted: 10 September 2018

Published online: 26 October 2018

References

- Laumann, T. O. et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron* **87**, 657–670 (2015).
- Poldrack, R. A. Precision neuroscience: dense sampling of individual brains. *Neuron* **95**, 727–729 (2017).
- Poldrack, R. A. et al. Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* **6**, 8885 (2015).
- Scarpazza, C. & De Simone, M. S. Voxel based morphometry: current perspectives. *Neurosci. Neuroeconomics* **5**, 19–35 (2016).
- Scarpazza, C. et al. When the single matters more than the group (II): addressing the problem of high false positive rates in single case voxel based morphometry using non-parametric statistics. *Front. Neurosci.* **10**, 6 (2016).
- Scarpazza, C., Sartori, G., De Simone, M. S. & Mechelli, A. When the single matters more than the group: very high false positive rates in single case Voxel Based Morphometry. *Neuroimage* **70**, 175–188 (2013).
- Farahany, N. A. Neuroscience and behavioral genetics in US criminal law: an empirical analysis. *J. Law. Biosci.* **2**, 485–509 (2015).
- Jones, O. D., Wagner, A. D., Faigman, D. L. & Raichle, M. E. Neuroscientists in court. *Nat. Rev. Neurosci.* **14**, 730–736 (2013).
- McSwiggan, S., Elger, B. & Appelbaum, P. S. The forensic use of behavioral genetics in criminal proceedings: case of the MAOA-L genotype. *Int. J. Law Psychiatry* **50**, 17–23 (2017).
- Meynen, G. A. A neurolaw perspective on psychiatric assessments of criminal responsibility: decision-making, mental disorder, and the brain. *Int. J. Law Psychiatry* **36**, 93–99 (2013).
- Sartori, G., Pellegrini, S. & Mechelli, A. Forensic neurosciences: from basic research to applications and pitfalls. *Curr. Opin. Neurol.* **24**, 371–377 (2011).
- Scarpazza, C., Pellegrini, S., Pietrini, P. & Sartori, G. The role of neuro-science in the evaluation of mental insanity: on the controversies in Italy. Comment on "on the stand. Another episode of neuroscience and law discussion from Italy". *Neuroethics* **11**, 83–95 (2018).
- Maoz, U. & Yaffe, G. What does recent neuroscience tell us about criminal responsibility? *J. Law. Biosci.* **3**, 120–139 (2016).
- Fenwick, P. Automatism, medicine and the law. *Psychol. Med. Monogr. Suppl.* **17**, 1–27 (1990).
- Ciccione, J. R. & Ferracuti, S. Comparative forensic psychiatry: II. The perizia and the role of the forensic psychiatrist in the Italian legal system. *Bull. Am. Acad. Psychiatry Law.* **23**, 453–466 (1995).
- Dror, I. E. Biases in forensic experts. *Science* **360**, 243 (2018).
- Murrie, D. C., Boccaccini, M. T., Guarnera, L. A. & Rufino, K. A. Are forensic experts biased by the side that retained them? *Psychol. Sci.* **24**, 1889–1897 (2013).
- Dror, I. E. Cognitive neuroscience in forensic science: understanding and utilizing the human element. *Phil. Trans. R. Soc. B* **370**, 20140255 (2015).
- Dror, I. E. Human expert performance in forensic decision making: seven different sources of bias. *Aust. J. Forensic Sci.* **49**, 541–547 (2017).
- Zapf, P. A. & Dror, I. E. Understanding and mitigating bias in forensic evaluation: lessons from forensic science. *Int. J. Forensic Ment. Health* **16**, 227–238 (2017).
- Zapf, P. A., Kukucka, J., Kassir, S. M. & Dror, I. E. Cognitive bias in forensic mental health assessment: Evaluator beliefs about its nature and scope. *Psychol., Public Policy, Law.* **14**, 1–10 (2018).
- Dror, I. E. A hierarchy of expert performance. *J. Appl. Res. Mem. Cogn.* **5**, 121–127 (2016).
- Dror, I. E. & Murrie, D. C. A Hierarchy of expert performance applied to forensic psychological assessment. *Psychol., Public Policy Law.* **24**, 11–23 (2018).
- Rosenhan, D. L. On being sane in insane places. *Science* **179**, 250–258 (1973).
- Gowensmith, W. N., Murrie, D. C. & Boccaccini, M. T. How reliable are forensic evaluations of legal sanity? *Law Hum. Behav.* **37**, 98–106 (2013).
- Miller, P. R. Inpatient diagnostic assessments: 2. Interrater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Res.* **105**, 265–271 (2001).
- Miller, P. R., Dasher, R., Collins, R., Griffiths, P. & Brown, F. Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. *Psychiatry Res.* **105**, 255–264 (2001).
- Guarnera, L. A., Murrie, D. C. & Boccaccini, M. T. Why do forensic experts disagree? Sources of unreliability and bias in forensic psychology evaluations. *Transl. Issues Psychol. Sci.* **3**, 143–152 (2017).
- Murrie, D. C. & Warren, J. I. Clinicians variation in rates of legal sanity options: implications for self-monitoring. *Prof. Psychol. Res. Pract.* **36**, 519–524 (2005).
- Hughes, V. Science in court: head case. *Nature* **464**, 340–342 (2010).
- Rigoni, D. et al. How neuroscience and behavioral genetics improve psychiatric assessment: report on a violent murder case. *Front. Behav. Neurosci.* **4**, 160 (2010).
- Fozdar, M. A. The relevance of modern neuroscience to forensic psychiatry practice. *J. Am. Acad. Psychiatry Law* **44**, 145–150 (2016).
- Reeves, D., Mills, M. J., Billick, S. B. & Brodie, J. D. Limitations of brain imaging in forensic psychiatry. *J. Am. Acad. Psychiatry Law* **31**, 89–96 (2003).

34. Husted, D. S., Myers, W. C. & Lui, Y. The limited role of neuroimaging in determining criminal liability: an overview and case report. *Forensic Sci. Int.* **179**, e9–e15 (2008).
35. Ashburner, J. & Friston, K. J. Voxel-based morphometry—the methods. *Neuroimage* **11**(6 Pt 1), 805–821 (2000).
36. Mechelli, A., Price, C. J., Friston, K. J. & Ashburner, J. Voxel-based morphometry of the human brain: methods and applications. *Curr. Med. Imaging Rev.* **1**, 00–00 (2005).
37. Roskies, A. L. Neuroimaging and inferential distance. *Neuroethics* **1**, 19–30 (2008).
38. Roskies, A. L., Schweitzer, N. J. & Saks, M. J. Neuroimages in court: less biasing than feared. *Trends Cogn. Sci.* **17**, 99–101 (2013).
39. McCabe, D. P. & Castel, A. D. Seeing is believing: the effect of brain images on judgments of scientific reasoning. *Cognition* **107**, 343–352 (2008).
40. Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E. & Gray, J. R. The seductive allure of neuroscience explanations. *J. Cogn. Neurosci.* **20**, 470–477 (2008).
41. Schweitzer, N. J. & Saks, M. J. Neuroimage evidence and the insanity defense. *Behav. Sci. Law* **29**, 592–607 (2011).
42. Dror, I. E., McCormack, B. & Epstein, J. Better science for a better justice: a proposal from joint experts. *Sci. Justice* (2018).
43. Fornari, U. & Galliani, I. Il caso giudiziario di Gianfranco Stevanin. *Centro Scientifico Editore, Torino*. (2003).
44. Stroop, J. R. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662 (1935).
45. Dubois, B., Slachevsky, A., Litvan, I. & Pillon, B. The FAB: a Frontal Assessment Battery at bedside. *Neurology* **55**, 1621–1626 (2000).
46. Simpson, G., Blaszczyński, A. & Hodgkinson, A. Sex offending as a psychosocial sequela of traumatic brain injury. *J. Head. Trauma. Rehabil.* **14**, 567–580 (1999).
47. Simpson, G., Tate, R., Ferry, K., Hodgkinson, A. & Blaszczyński, A. Social, neuroradiologic, medical, and neuropsychologic correlates of sexually aberrant behavior after traumatic brain injury: a controlled study. *J. Head. Trauma. Rehabil.* **16**, 556–672 (2001).
48. Aron, A. R., Robbins, T. W. & Poldrack, R. A. *Inhib. Right Infer. Front. cortex: One Decade on.* *Trends Cogn. Sci.* **18**, 177–185 (2014).
49. Bari, A. & Robbins, T. W. Inhibition and impulsivity: behavioral and neural basis of response control. *Prog. Neurobiol.* **108**, 44–79 (2013).
50. Chamberlain, S. R. & Sahakian, B. J. The neuropsychiatry of impulsivity. *Curr. Opin. Psychiatry* **20**, 255–261 (2007).
51. Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50**, 7–15 (1994).
52. Roma, P., Pazzelli, F., Pompili, M., Girardi, P. & Ferracuti, S. Shibari: double hanging during consensual sexual asphyxia. *Arch. Sex. Behav.* **42**, 895–900 (2013).
53. Burgess, P. W. & Shallice, T. Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia* **34**, 263–272 (1996).
54. Shallice, T. Specific impairments of planning. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **298**, 199–209 (1982).
55. Blair, R. J. & Cipolletti, L. Impaired social response reversal. A case 'acquired sociopathy'. *Brain*. **123**(Pt 6), 1122–1141 (2000).
56. Prior, M., Sartori, G. & Marchi, S. Cognizione sociale e comportamento. *UPSEL Domeneghini*. (2003).
57. Cho, S. S. et al. Morphometric correlation of impulsivity in medial prefrontal cortex. *Brain Topogr.* **26**, 479–487 (2013).
58. Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J. & Robbins, T. W. Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat. Neurosci.* **6**, 115–116 (2003).
59. Li, C. S., Huang, C., Constable, R. T. & Sinha, R. Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. *J. Neurosci.* **26**, 186–192 (2006).
60. Paradiso, S. et al. Cerebral blood flow changes associated with attribution of emotional valence to pleasant, unpleasant, and neutral visual stimuli in a PET study of normal subjects. *Am. J. Psychiatry* **156**, 1618–1629 (1999).
61. O'Neill, A. et al. Magnetic resonance imaging in patients with borderline personality disorder: a study of volumetric abnormalities. *Psychiatry Res.* **213**, 1–10 (2013).
62. Zhao, Y. J. et al. Brain grey matter abnormalities in medication-free patients with major depressive disorder: a meta-analysis. *Psychol. Med.* **44**, 2927–2937 (2014).
63. Moreno-Lopez, L. et al. Trait impulsivity and prefrontal gray matter reductions in cocaine dependent individuals. *Drug Alcohol. Depend.* **125**, 208–214 (2012).
64. Doallo, S. et al. Larger mid-dorsolateral prefrontal gray matter volume in young binge drinkers revealed by voxel-based morphometry. *PLoS One* **9**, e96380 (2014).
65. Poldrack, R. A. Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* **10**, 59–63 (2006).
66. Poldrack, R. A. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron Dec.* **72**, 692–697 (2011).
67. Choi, O. S. What neuroscience can and cannot answer. *J. Am. Acad. Psychiatry Law* **45**, 278–285 (2017).
68. Sartori, G., Scarpazza, C., Codognotto, S. & Pietrini, P. An unusual case of acquired pedophilic behavior following compression of orbitofrontal cortex and hypothalamus by a Clivus Chordoma. *J. Neurol.* **263**, 1454–1455 (2016).
69. Farisco, M. & Petrini, C. On the stand. Another episode of neuroscience and law discussion from Italy. *Neuroethics* **7**, 243–245 (2014).
70. Rojas-Burke, J. PET scans advance as tool in insanity defense. *J. Nucl. Med.* **34**, 13N–16N (1993).
71. Ferracuti, S., Accornero, N. & Manfredi, M. Atypical psychosis associated with left temporal arachnoid cyst: report of four cases. *Integr. Psychiatry* **7**, 132–139 (1991).
72. Lockhart, J. J. & Satya-Murti, S. Diagnosing crime and diagnosing disease: bias reduction strategies in the forensic and clinical sciences. *J. Forensic Sci.* **62**, 1534–1541 (2017).
73. Darby, R. R., Horn, A., Cushman, F. & Fox, M. D. Lesion network localization of criminal behavior. *Proc. Natl Acad. Sci. USA* **115**, 601–606 (2018).
74. Adolphs, R., Glascher, J. & Tranel, D. Searching for the neural causes of criminal behavior. *Proc. Natl Acad. Sci. USA* **115**, 451–452 (2018).
75. Bogousslavsky, J. Artistic creativity, style and brain disorders. *Eur. Neurol.* **54**, 103–111 (2005).
76. Granata, T. et al. Hemispherotomy in Rasmussen encephalitis: long-term outcome in an Italian series of 16 patients. *Epilepsy Res.* **108**, 1106–1119 (2014).
77. Denney, R. L. Symptom Validity Testing of remote memory in a criminal forensic setting. *Arch. Clin. Neuropsychol.* **11**, 589–603 (1996).
78. Silva, J. A. Forensic psychiatry, neuroscience, and the law. *J. Am. Acad. Psychiatry Law* **37**, 489–502 (2009).
79. Radua, J. et al. Multimodal meta-analysis of structural and functional brain changes in first episode psychosis and the effects of antipsychotic medication. *Neurosci. Biobehav. Rev.* **36**, 2325–2333 (2012).
80. Mesulam, M. The evolving landscape of human cortical connectivity: facts and inferences. *Neuroimage* **62**, 2182–2189 (2012).
81. Mesulam, M. M. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* **28**, 597–613 (1990).
82. Glascher, J. et al. Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proc. Natl Acad. Sci. USA* **109**, 14681–14686 (2012).