CHAPTER SEVEN

Neuroscientific approaches to 'mens rea' assessment

Ullrich Wagner and Henrik Walter

INTRODUCTION

'Mens rea' (Latin for 'guilty mind') is a central concept in current legal systems of Western countries. There it is part of the standard common law test of criminal liability, as expressed in the Latin phrase: 'Actus non facit reum nisi mens sit rea.' (An act does not make someone guilty unless the mind is likewise guilty.) Thus, a 'guilty mind' is a necessary element of a crime, i.e. a culprit can only be legally blamed for a criminal act (the 'actus reus') if this act was committed deliberately. Accordingly, the legal system not only has to prove that it was actually the accused person who performed the criminal act. Beyond this task, much of the work of juries and judges refers to determining the beliefs, intentions, and desires of the culprit that made his mind 'guilty' at the time of the crime.

The present chapter examines what neuroscience can contribute to this legal process of 'mens rea' assessment, summarizing the current state of relevant empirical findings in cognitive neuroscience. As such, it aims to contribute to an emerging new field within social neuroscience, called 'neurolaw', trying to connect neuroscience and law (Goodenough & Prehn, 2004; Gazzaniga, 2008; Schleim, Spranger, & Walter, 2009; Goodenough & Tucker, 2010; Müller & Walter, 2011). Although few neuroscientific studies have directly addressed legal topics so far, social neuroscience has meanwhile investigated a variety of cognitive processes pertinent to the legal process of 'mens rea' assessment, such as belief attribution, moral judgement, and deception. Therefore, social neuroscience could provide useful additional information on how a 'mens rea' can be identified.

We are aware of the difficult relationship between law and neuroscience due to different concepts, aims, and traditions in the two fields, which became apparent in the controversial debate on the existence and the role of 'free will' in human decisions (Walter, 2001, 2011; Roskies, 2006). By focusing on the more specific and practically relevant question of 'mens rea' assessment, we hope to circumvent these basic discussions that frequently end up in an impasse without a real mutual exchange. In this way, we want to open the view for the insights that neuroscience may actually offer with regard to processes that take place in the legal system every day, with the ultimate aim to help to improve them.

We will first focus on brain processes of the culprit that might indicate 'mens rea'. Then, we will summarize cognitive neuroscience studies in brain processes of subjects in the role of judges or jurors who evaluate 'mens rea' in a culprit. Finally, we will discuss how these findings can impact on legal theorizing and legal practice.

NEUROSCIENTIFIC ASSESSMENT OF 'MENS REA' INDICATORS IN THE CULPRIT'S BRAIN

The heart of the legal process in criminal law is the collection of facts that are relevant to possible sentencing of the accused. 'Mens rea' can be regarded as one of these facts to be collected. From its beginning, law had to struggle with the problem that there was no direct way to look into the culprit's mind, so 'mens rea' had always been determined by judges or jurors indirectly by evaluating other facts and testimonies, which can be subject to mistakes and misjudgements. Even if a culprit confesses the crime, this is not per se an ultimate proof. A theoretical ideal for the legal process would therefore be a procedure that would be capable of assessing a 'guilty mind' of a culprit objectively, as an undisputable fact – similar to fingerprints or DNA traces that can be seen as a 'proof beyond a reasonable doubt' that the accused was actually present at the scene of the crime. Can neuroscience help here? At least, neuroscientific methods have meanwhile reached a level that allows for predicting a person's intentions with a certain accuracy.[1] This has been demonstrated in a study by Haynes et al. (2007). In this study, subjects undergoing functional magnetic resonance imaging (fMRI) scanning were shown two two-digit numbers and were asked in each trial – during a preparation phase that immediately preceded the presentation of the numbers – to freely choose to either add or subtract (addition and subtraction intention) the second from the first number. Using a method called multivoxel pattern analysis, the authors were able to predict with 71% accuracy from patterns of activations within the anterior medial prefrontal cortex (MPFC) during the preparation phase which of the two tasks the subject subsequently performed. In other words, to some degree the intentions of the subject could be identified from their brain activation. Although these are impressive initial findings, the decoding accuracy of 71%, compared to 50% chance level, is certainly not yet sufficient to call this method 'mind reading' as most people would understand it. But even with substantial improvement of the decoding accuracy that might be possible in the future it is unlikely that such procedures could ever be actually applied in the legal system to determine 'mens rea', for at least two reasons. First, compared to real-life situations, the experimental setting was very restricted, with only two pre-defined possible intentions that the subjects were allowed to choose from, and subjects had to be, and actually were, cooperative. Second, and theoretically even more important, the law requires the existence of 'mens rea' at the moment of the crime. A direct assessment would only be possible in the unrealistic scenario in which brain activity in the culprit could be measured online *during* the criminal act.

Thus, even with the application of neuroscientific methods, however exact and reliable they might be in the future, determining 'mens rea' will always remain an indirect post-hoc procedure based on facts collected after the crime. But neuroscience may also be useful in such indirect post-hoc methods to determine guilty minds. In fact, this field of fact collection was the first one where neuroscientific methods have been actually applied in the legal system, namely by use of biological measures of 'lie detection' (Segrave, 2004; Grubin, 2010). The rationale is that an actual perpetrator who claims to be not guilty needs to hide at trial certain knowledge that he has about the crime. If a neuroscientific device were capable of detecting lies as opposed to truthful statements, his objectively measured lying or deception (e.g. when he states that he does not know the kind of weapon that was used in a certain crime) could be used as evidence in the trial.[2] Since the early 20th century, the so-called 'polygraph', a device basically relying on combined peripheral measurements of skin conductance, heart rate, and respiration as an indicator of physiological arousal, has actually been applied as such a lie detector in legal contexts, predominantly in the USA (Grubin, 2010).[3] Undergoing careful scientific scrutiny, however, the reliability of the polygraph as a lie detector turned out to be insufficient, so that the US Supreme Court finally barred this method in court-martial proceedings (U.S. *vs.* Scheffer, 1998),[4] which was similarly decided in other countries, e.g. Germany.[5]

However, neuroscience is a quickly developing scientific field, so the question arises whether other neuroscientific methods could replace the polygraph as lie detectors. Meanwhile, several studies have been performed using fMRI to detect deception (e.g. Kozel et al., 2005; Langleben et al., 2005; Mohamed et al., 2006; for overviews, see Sip et al., 2008; Seiterle, 2010). Although quite different experimental paradigms have been applied (including the 'comparison question test' and the 'guilty knowledge test' that are also widely used in polygraph applications), several brain areas related to cognitive control and attention, including dorsolateral prefrontal cortex (DLPFC) and anterior cingulate cortex (ACC), have been shown to be activated more strongly during lying than during truth-telling (Sip et al., 2008). This is consistent with the cognitively demanding requirements of information management and impression management, encompassing inhibition of the prepotent true answer and monitoring the reactions of the communication partner. These are interesting findings from a scientific point of view, but at the present stage, application of fMRI research in this field is still in its infancy and, like the polygraph, is presently not considered to fulfil the standards required for its use as a 'lie detector' (Miller, 2010; Schauer, 2010; Seiterle, 2011). In the future, the use of multivoxel pattern analysis, as in the attempts of direct decoding intentions described above, may improve this method with regard to the predictive value in individuals (Davatzikos et al., 2005). However, there are also more general methodological problems with these 'lie detection' studies using fMRI, apart from purely technical questions and questions of statistical analysis. For example, in the typical experimental paradigms of deception used so far in fMRI studies, subjects are *instructed* to deceive, which eliminates the aspect of voluntary intention that is a characteristic feature of real-life deception (Sip et al., 2008). Also, there is little at stake for subjects in a laboratory experiment, but a real liar typically has much to lose, particularly in legal contexts. The scientific progress in this field is not made easier by the fact that meanwhile some researchers who are particularly active in the field also have financial interests due to their involvement in 'truth verification companies' like Cephos and No Lie MRI, which counteracts their scientific independence and also constrains the use of such methods by other independent researchers due to patents that these companies hold on certain procedures.

On the other hand, it has been argued that the usefulness of the application of neuroscientific tools does not require that they meet all scientific standards in order to be acceptable according to legal standards (Schauer, 2010; Seiterle, 2011). A recent decision by a court in the US state of Tennessee confirms this (U.S. *vs.* Semrau, 2010; cf. Miller, 2010; Seiterle, 2011). The judge had to decide whether or not to admit presentation of fMRI data from the defendant's brain. The question was if the defendant Lorne Semrau, accused of fraud against several health-care providers, acted intentionally when causing the financial harm to these companies. The defence claimed that he did not and introduced fMRI evidence provided by the company 'Cephos' for this claim. This case is interesting, as this was the first so-called 'Daubert' hearing for fMRI lie detection. The 'Daubert rules' were established by the US Supreme Court in 1993 as guidelines to weigh the admissibility of scientific evidence in courts in the USA (Daubert *vs.* Merrell Dow Pharmaceuticals, 1993). These guidelines suggest – although do not absolutely demand – that a given technology should meet four criteria: it should be subject to empirical testing, be published in peer-reviewed literature, have a known error rate, and be generally accepted in the scientific community (Miller, 2010; Seiterle, 2011). The judge's report came to the conclusion that fMRI lie detection currently meets only the first two criteria. Nevertheless, it left the door open for new decisions on this issue in the future. One of the reasons for his decision was the lack of studies outside the lab situation, so that information on error rates for real-life situations was missing. Critically, he explicitly added that the existence of such information would not be a *necessary* condition for admissibility of fMRI technology, which could be admitted even in the absence of such information in the future if the methods and procedures improve otherwise (Seiterle, 2011). If this argument prevails, the actual use of fMRI 'lie detection' technology as standard evidence in courtrooms may come sooner, as many legal scholars currently would expect.

Up to now we have focused on the question of how neuroscientific tools could be used to provide evidence for the presence of 'mens rea' in a certain individual accused of having committed a specific crime. However, neuroscientific methods can likewise be used in legal defence of the accused by providing evidence for the *absence* of 'mens rea'. Here, neuroscience may identify anomalies in brain structure or function in the accused individual that would exclude intentional and wilful acting that is implied by 'mens rea'. Such evidence can then be part of the so-called 'insanity defence', which – according to the M'Naghten rule[6] used as a standard in most Anglo-American jurisdictions – assumes a 'defect of reason, from disease of the mind' in the culprit, which makes him unable 'to know the nature and quality of the act he was doing; or if he did know, that he did not know he was doing what was wrong' (Wettstein, Mulvey, & Rogers, 1991). In contrast to the more narrow sense of 'mens rea' determination in mentally healthy people (or 'reasonable persons', in legal terms), such evidence usually indicates a more general mental distortion in the respective individual that would likely also affect other life situations so that the individual is regarded as unable to stand trial at all.[7] Schizophrenia, which can be accompanied by severe hallucinations, is an example of a mental disease that can lead to this type of legal defence. Most likely, M'Naghten himself, who shot the British Prime Minister's secretary, and whose case originally led in the 19th century to the formulation of the M'Naghten rule mentioned above, suffered from this mental disorder (Bennett, 2009). The classical diagnostic tools of forensic psychiatry that usually do not include neuroscientific examinations are mostly sufficient to determine such circumstances that result in the verdict of 'not guilty by reason of insanity' (NGRI). However, with the advent of more and more sophisticated neuroscientific diagnostic tools, data from these technologies have also meanwhile found their way into the courtrooms for this purpose (Moriarty, 2008). One prominent example was John Hinckley, who attempted to assassinate President Ronald Reagan in 1981. His lawyers successfully introduced neuroimaging evidence at trial as proof of his insanity, specifically presenting a computer-assisted tomography (CAT) scan that showed widened sulci in his brain that are typical for (although not specific to) schizophrenic patients (Batts, 2009).

Also 'automatism' can be used as a legal defence, which means that the criminal act was not really an 'action' at all (in the legal sense of a willed action). This can be the case in movements resulting from epileptic seizures or from certain sleep disturbances. For example, a man who killed his wife during sleep was recently acquitted in the UK because he could be shown to suffer from a sleep disorder that led him to incorporate his wife as a dangerous enemy into a nightmare (de Bruxelles, 2009). Careful examination by neuroscientifically educated experts is usually required in such cases, showing that neuroscience can provide useful services in these legal decisions. Likewise, many European countries also include the 'lack of control of one's behaviour' as a criterion in the insanity defence. Notably, in the USA this 'volitional test' of the insanity defence, i.e. lack of control, was eliminated as a legal insanity standard on the federal level after the case of the Reagan assassin (Schopp, 1991).[8]

Apart from these extreme and relatively rare cases of insanity, neuroscientific evidence can also be used legally in most countries to indicate 'diminished responsibility' that can justify reduced penalties. In the USA, although the U.S. Model Penal Code (American Law Institute, 1962) defines different modes of culpability which reflect different levels of 'mens rea' (specifically distinguishing whether the act was committed purposefully, knowingly, recklessly, or negligently), the definition of 'diminished responsibility' remains unspecified. Theoretically, any kind of evidence that could be seen as a sign of impaired mental responsibility can therefore be incorporated, which also leaves space for the consideration of a variety of brain abnormalities as mitigating factors. In fact, as pointed out by Moriarty (2008), courts – although still generally sceptical against neuroimages as legal proof – have admitted neuroscientific evidence most willingly in the penalty phase of a trial, particularly in death penalty cases, where such evidence in some cases led to a change to life sentence (Batts, 2009).[9]

A general question is how population-based neuroscientific findings can (and should) be used as evidence in the legal context of 'mens rea' determination. The legal system inherently deals with single cases, but neuroscientific studies, in order to be able to generalize conclusions, usually provide aggregated data from subject groups that are representative of a whole population of individuals, e.g. the population of mentally healthy people, the population of schizophrenic patients, or the population of antisocial/violent individuals. Findings regarding the last group are particularly relevant here because individuals from this group are tremendously over-represented in the legal context. It is estimated that up to 75% of all prison inmates fulfil the DSM-IV criteria of antisocial personality disorder (Weber et al., 2008). About a quarter of these prisoners additionally show the diagnostic criteria of psychopathy, which include personality traits of emotional detachment such as callousness, manipulativeness, and lack of remorse and empathy (Hare, 1991). A variety of neuroscientific investigations have shown that antisocial and psychopathic persons show a number of structural and functional abnormalities in brain regions, predominantly in the prefrontal and temporal lobes, that are involved (in mentally healthy people) in the emotional processing and in moral judgement (Raine & Yang, 2006; Walter et al., 2009; Yang & Raine, 2009).[10] One of the areas most consistently reported to be compromised in these populations is the orbitofrontal cortex (OFC). Interestingly, exactly this area is specifically activated in normal subjects when they experience the feeling of guilt (Wagner et al., 2011). Wagner and colleagues performed a study in which different social emotions (guilt, shame, and sadness) were induced by reliving specific autobiographical memories from the past. Activity in the right OFC was stronger in guilt than in the closely related emotions shame and sadness and also correlated across subjects with individual propensities to experience guilt ('trait guilt'). Interestingly, if (formerly) healthy people suffer from brain lesions in this area, e.g. due to an accident, they frequently also show signs of antisocial behaviour germane to psychopathy (so-called 'acquired sociopathy'; Blair, 2001), which may result from the lack of behavioural control that is normally accomplished by the ability to experience anticipated feelings of guilt.

While psychopaths appear to lack the ability to experience certain socially relevant emotions like guilt and therefore tend to show 'instrumental' (or 'cold') aggressiveness, other subgroups of antisocial individuals who tend to show more 'impulsive' (or 'hot') aggressiveness seem to lack the ability to *regulate* their emotions appropriately. In healthy people this ability involves the right DLPFC and the right parietal cortex, which affect the amygdala via the orbitofrontal cortex (Walter et al., 2009; Erk et al., 2010). Reduced metabolism of the DLPFC has been found in different groups of aggressive subjects (Hirono et al., 2000; Juhasz et al., 2001).

These few examples may be sufficient to demonstrate that there are neurobiological factors that, at least in a probabilistic manner, underlie proneness to certain types of aggressive behaviour that is likely to result sooner or later in legal prosecution. The critical term here is 'probabilistic', which does not allow a reliable statement on a specific individual. But reliable statements about a person are generally the exception rather than the rule in legal trials, so this would not be an argument per se against the use of such neuroscientific information. We will come back to this issue in the concluding part of this chapter.

HOW THE JUDGE'S BRAIN EVALUATES 'MENS REA' OF A CULPRIT

Although the use of neuroscientific evidence in courts will probably be extended considerably in the future, it has also become clear from the previous section that due to both practical and theoretical constraints, determination of 'mens rea' in a culprit will certainly never rely exclusively on direct neuroscientific assessment of brain activity in that person. Thus, it will ultimately remain the responsibility of human actors in the legal system (judges and juries) to accomplish this task. But the actions and decisions of these human beings likewise rely on neurobiological processes in their brains. Thus, another neuroscientific approach to contribute to the understanding of 'mens rea' assessment is to observe the brains of people who are judging what other people (in the legal context typically the defendant accused to have committed a crime) think, believe, or desire. This psychological process is called 'mentalizing' or, more frequently, 'Theory of Mind' (ToM), a term initially introduced by Premack and Woodruff (1978) in the context of the question of whether apes can correctly understand what goes on in the minds of their conspecifics. Although this is the case to a certain degree, only humans have developed ToM capabilities to such an extent that made the complex social life in our modern societies (including their legal systems) possible. In fact, sophisticated ToM capabilities, together with the use of language, can be regarded as the critical factor that makes humans unique among the animals (Brüne & Brüne-Cohrs, 2006; Saxe, 2006). In the present context, the term 'Theory of Mind' can be taken quite literally, because judges and jurors are using different pieces of evidence to form a theory of the state of the mind of the defendant when he committed a criminal act, in order to determine whether it was a 'guilty' mind.

Initially, most researchers investigating ToM capabilities in humans were specifically interested in the ontogeny of these capabilities in children. Here, the typical experimental paradigm used for this purpose was the 'false belief' task originally introduced by Wimmer and Perner (1983; see also Wellman, Cross, & Watson, 2001 for an overview). In this task, the child under investigation is shown a picture story in which a character's belief about a location of an object becomes wrong when the object is moved without the character's knowledge. For example, in the prototypical 'Sally–Anne test', Sally puts a ball into one of two boxes. After Sally has left the room, Anne moves the ball into the other box. Then Sally comes back. The task of the child is to say where Sally will look for the ball upon her return. The critical feature in such 'false belief' tasks is that it requires the child to distinguish between his/her own knowledge about reality and another person's beliefs, an 'acid test' (Frith & Frith, 1999) of ToM. As the developmental studies have shown, this capability emerges at the age of 3–4 years old. Before this age, children do not use beliefs to explain the actions of other persons.

Neuroscientific studies aiming at revealing the neural underpinnings of ToM first applied similar versions of the 'false belief' paradigm to adult subjects (Fletcher et al., 1995; Gallagher et al., 2000; Vogeley et al., 2001; Saxe & Kanwisher, 2003). These studies identified the MPFC, the temporal poles, the superior temporal sulcus (STS), predominantly in posterior parts, and the adjacent temporo-parietal junction (TPJ) as critical brain regions involved in ToM. Interestingly, although later studies have investigated ToM capabilities in a variety of different ToM paradigms apart from the 'simple' false belief task, the same areas can still be regarded as the 'core' ToM network, together with the precuneus and posterior cingulate cortex (and, less consistently, the amygdala) which are typically activated as well across different tasks (Carrington & Bailey, 2009).

Further studies have shown that the ToM network gets differently involved depending on the type of intentions involved. While representations of simple intentions ('wanting to read') only activate the right TPJ and the posterior cingulate, the anterior parts of the network as well as the left TPJ become specifically active for communicative intentions (for example, 'A signals B that he wants to drink something') (Walter et al., 2004; Ciaramidaro et al., 2007). The right TPJ appears to be most specifically implicated in basic belief attribution. As demonstrated by Saxe and Kanwisher (2003), this brain region shows increased responses not only in false belief tasks but generally in tasks that invite ToM reasoning about another person's beliefs, regardless of whether they are true or false. As a test of specificity, these authors further showed that the same area does not respond to unspecific social processing, i.e. the mere presence of a person without attribution of beliefs to that person.

For the purpose of the present chapter, ToM reasoning about beliefs becomes particularly important when the behaviour of a person has to be evaluated morally (see also de Oliveira-Souza, Zahn, & Moll, this volume).[11] In the context of developmental research, the term 'morally relevant theory of mind' (MoToM) has recently been coined for such situations (Killen et al., 2011). It has been well known since the classical studies in developmental moral psychology performed by Jean Piaget in the early 20th century that young children's moral judgements rely solely on the outcome of a person's actions, while older children increasingly also take the actor's beliefs and intentions into account. For example, judging either a person who intends to direct a traveller to the right location but accidentally misdirects him or a person who intends to misdirect the traveller but accidentally leads him to the right place, younger children consider the former, but older children the latter, as more blameworthy (Piaget, 1932). Exactly this more mature moral judgement performed by older children and adults, focusing more on what goes on in the actor's mind, is reflected in the legal differentiation between 'actus reus' and 'mens rea'.

What are the neural bases of this integration of outcome and beliefs in moral judgements? A number of neuroscientific studies performed by Rebecca Saxe and her colleagues again point to a critical involvement of the right TPJ in this process (Young et al., 2007; Young & Saxe, 2009). In these studies, they used an experimental paradigm in which the two critical factors 'outcome' and 'belief' were varied independently in written scenarios, where the behaviour of a protagonist had to be judged by their subjects. For example, in one of these scenarios a protagonist puts white powder into the coffee of a colleague, which is either sugar (neutral outcome) or a toxic substance that leads to the death of the colleague (negative outcome), and the protagonist believes either that the powder is sugar (neutral belief) or that it is toxic (negative belief). The two critical conditions here are those where belief and actual outcome do not coincide, especially in the case of an attempted harm, where 'mens rea' is present without actually achieving the intended effect (crime of attempt). In fact, subjects judged this condition of intended harm without effect as blameworthy as intended harm with actually achieving the intended effect. In contrast, moral blameworthiness of negative outcome was substantially lower when they were unintended (unknowing harm) than when they were intended. Critically, on the brain level, this interaction between beliefs and outcome was specifically reflected in the activation of the right TPJ (and to some degree also in the dorsal part of the MPFC) amongst the regions of interest related to ToM (Young et al., 2007). In support of these results, Young and Saxe (2009) further found that individual differences in right TPJ activation predicted how much subjects were willing to reduce their moral blame for actors who committed accidental harm.

In a subsequent study, Young et al. (2010) additionally demonstrated that the right TPJ indeed plays a causal role in the integration of belief information in moral judgement. Using the same scenarios, they applied transcranial magnetic stimulation (TMS) to transiently disrupt neural activity of the right TPJ while subjects performed moral judgements on the protagonists. Disrupting the right TPJ in this way (but not disrupting a nearby control region in the parietal cortex) led to less harsh moral disapproval of attempted harm, while leaving judgements in the other conditions unaffected. Thus, interfering with neural processing in the right TPJ was specifically effective in the condition where 'mens rea' assessment plays the predominant role in determining the extent of moral blameworthiness. In terms of developmental psychological research described above, subjects were 'set back' to a young child's level of moral evaluation, taking predominantly outcome information into account.

These and other studies of moral judgement (see de Oliveira-Souza, Zahn, & Moll, this volume) are clearly relevant to the understanding of judgemental processes in the legal domain. However, they were not directly designed for this purpose. Only recently has research in social neuroscience also begun to directly model legal procedures experimentally to identify the underlying neural mechanisms. One such study addressed the neural mechanisms of third-party punishment, i.e. punishment by an uninvolved and impartial person, as the most distinctive feature of legal decision making in criminal law (Buckholtz et al., 2008). Subjects in this study, adopting the role of a judge, read written scenarios of crimes and had to assign the appropriate punishment to the described culprits. In some scenarios, there were mitigating circumstances (diminished responsibility), in others not (full responsibility). Regarding categorical analyses, the right DLPFC was more strongly activated in scenarios of full as compared to diminished responsibility and, within the scenarios of diminished responsibility, was more strongly activated when subjects decided to punish than when they decided not to punish. Thus, the right DLPFC appears to be critically involved in the basic decision of whether a culprit should be punished or not based on the assessment of criminal responsibility. However, in parametric analyses, the magnitude of punishment assigned was predicted by activity in other areas, namely amygdala, posterior cingulate, temporal pole, and MPFC, all of which are related to social-emotional processing and ToM (see above). Interestingly, the TPJ was not predictive of punishment in this way, but it was activated in the reverse contrast in the categorical analysis, i.e. it showed more activity in scenarios of diminished than full responsibility. Thus, consistent with the findings described above, the TPJ appears to come into play when subjects determine overall on the basis of mitigating circumstances whether 'mens rea' is present or not in the defendant, while other aspects of ToM may become relevant when, provided that 'mens rea' is regarded as present, the extent of punishment is determined.

Drawing on the observation in actual sentencing practice in the USA that repeat offenders commonly receive more severe punishments than first-time offenders, Kliemann et al. (2008) designed an experimental model to investigate the neural basis of this phenomenon in ordinary people's moral intuitions. They hypothesized that negative prior record would lead subjects to attribute more intentionality to agents causing negative outcomes, mediated by activation within the ToM network. Subjects read vignettes about an agent's action with subsequent positive or negative outcome, leaving the agent's mental states (intentions, beliefs, goals) open. Before this vignette task, to manipulate the perceived 'prior record' of the agent, subjects initially – outside the fMRI scanner – played an economic trust game with other players (presented with forename and photograph), which were purportedly the real actors later described in the vignettes. (To maintain the credibility of the procedure, at the beginning of the experiment each subject had to provide a photograph of him/herself and short descriptions of one event with a positive outcome and one event with a negative outcome from his/her personal past, which allegedly were transcribed into a vignette later presented to other subjects.) In the economic game, half of the (purported) co-players played fairly, the other half unfairly. Critically, in the subsequent vignette task, subjects indeed attributed more intentionality and gave more blame to players with a negative record (who had previously played unfairly) than to players with a positive record (who had previously played fairly), particularly in the case of negative outcomes. As expected, this interaction effect was reflected on the neural level in the right TPJ (and to a lesser degree also in other ToM regions such as precuneus, dorsal MPFC, and left TPJ), whose activation specifically increased when subjects judged the vignette with negative outcome of a previously unfair player. Thus, the legal practice of harsher punishment for repeat offenders appears to rely on intuitively enhanced moral blame due to enhanced attribution of intentionality to such offenders.

Only one neuroimaging study on legal issues so far has also investigated law experts (lawyers) rather than only the typical student population lacking specific legal education (Schleim et al., 2011). This is an important point, as expertise may change the way legal cases are neurally processed, especially against the background of the law ideal of purely rational judgement, free of emotion and passion, that is conveyed during legal education (Gewirtz, 1996). Further, such education may also enhance the difference between moral and legal judgement, compared to non-expert participants who can rely only on their (moral) intuitions when judging legal issues. To address these questions, Schleim and colleagues developed target stories on the basis of media reports and scholarly literature that were dilemmatic from a moral as well as a legal point of view. Two groups of subjects (20 lawyers and 20 legally unskilled other academics, matched for age, education, and gender) underwent fMRI scanning while judging the stories from a moral or a legal point of view. There were three main findings. First, across all subjects both moral and legal judgements commonly recruited a widespread 'moral brain' network (Moll & de Oliveira-Souza, 2007) encompassing, among others, critical ToM regions like anterior MPFC, posterior cingulate/precuneus, and TPJ. Second, legal judgement differed from moral judgement only in the decision phase, where the left DLPFC was more strongly activated, probably reflecting the stronger application of explicit rules. Third, an interaction with legal expertise (lawyers *vs.* legal laymen) was found only in the dorsal ACC, which was more strongly activated during legal judgement than moral judgement in lawyers, with a reverse pattern in legal laymen, suggesting an attention shift towards legal processing due to legal expertise. Notably, none of the typical ToM regions or typical regions of emotion processing was differentially activated in lawyers and legal laymen, despite the fact that, behaviourally, lawyers appeared to be less emotionally involved during judgement than the laymen. Together, the results show that, despite some differences, legal judgement basically recruits the same brain networks as moral judgement (with ToM regions as critical components), and this holds similarly for people with and without educational expertise in law.

IMPLICATIONS

In this chapter, we have described what neuroscience can contribute, practically and theoretically, in the legal task of 'mens rea' assessment. In the first part, we discussed how neuroscientific instruments could be directly applied to a certain defendant in order to determine his/her 'mens rea' in the context of a specific trial. Such tools could indeed be of practical use in certain legal contexts, and we have described examples where neuroscientific evidence has already been used in criminal law.[12] Even the decision by the US Supreme Court not to apply the death penalty to offenders below the age of 18 (Roper *vs.* Simmons, 2005) was partly influenced by the presentation of neuroscientific evidence of a still maturing brain at this age, with an ongoing increase of white matter specifically in the frontal lobe that is critically involved in impulse control (Beckman, 2004; Aronson, 2007). The time is over now for the possibility to just deny the existence of neuroscience or the claim that it is generally irrelevant in the legal domain. Even if the use of neuroscientific evidence is currently very limited (and is for many legal purposes indeed irrelevant or practically useless), its use will certainly increase in the future, and the mere fact that it *is* used in certain legal contexts makes it necessary to define standards for its use. As mentioned above, one of the problems to be solved here is how information that is only probabilistic in nature should be treated. For example, if a defendant belongs to a clinical group like schizophrenia or psychopathy, known to be associated with certain brain anomalies, which is certainly not alone sufficient as proof of an absent 'mens rea', what additional evidence is needed to regard him as 'non-guilty by reason of insanity'? Even in the absence of a clinical diagnosis, neuroimaging might in principle provide relevant information if there is an established standard. For example, if an accused person without signs of psychiatric disorders performs an emotion regulation task as used in healthy subjects and reduced activation or impaired connectivity in his DLPFC is demonstrated, what then is the standard of how much DLPFC activation is 'sufficient' or 'insufficient' to be able to inhibit impulsive aggression? If sufficient reliable and replicable data from different groups of healthy subjects have been collected, these could serve as comparison standards. However, a problem here lies in the fact that a culprit might not try hard enough, so that a missing brain process might not tell very much. On the other hand, in the context of preventive detention, e.g. in sex offenders, the question might be asked in the opposite direction, i.e. whether someone is able to control his impulses when confronted with a sexual stimulus or how much his cognitive control brain mechanism has been improved by therapy. This might be one piece of evidence in formulating a prognosis.

It is important to note that brain imaging does not principally reveal only *deficits* of internal control. A recent neuroimaging study has shown that the genetically determined emotional sensitivity of the amygdala in carriers of a certain variant of the serotonin transporter can be compensated by active cognitive emotion regulation strategies (Schardt et al., 2010). This is relevant information as, recently, a combination of genetic information (about a variant in another gene, which controls the synthesis of monoamine oxidase A [MAO-A]) and neuroimaging findings led to a reduction in sentencing for murder in an Italian case (Feresin, 2009).

As stated by Moriarty (2008), 'a neuroimage of a frontal lobe defect does not provide unequivocal proof of that person's lack of mens rea' (p. 47). That is true, but, at least in the future when neuroimaging techniques will have further improved and at least some rough standards will have been established,[13] it could provide another piece of (statistical) information to the mosaic of other (likewise mostly statistical) information, which altogether may at least in certain cases justify reduced penalties due to 'diminished responsibility'. As with accepting neuroimaging evidence in general, courts will probably be open to including such information when the death penalty is at stake. In principle, this would be just a further sophistication of what has already been accepted now by several courts.

It is important to note that all these questions are not specific to neuroscientific evidence, because courts also regularly rely on other information that is similarly subject to considerable uncertainty, such as eyewitness testimony (Busey & Loftus, 2007; Loftus, Doyle, & Dysert, 2008). Thus, the more general underlying question is how different pieces of evidence (which may or may not result from neuroscientific investigations), each of which is uncertain to some degree, can be combined to justify a reasonable legal decision. It is desirable that the legal system becomes more explicit in this regard. Those legal scholars who are sceptical specifically against the use of neuroscientific methods should first scrutinize with the same scepticism the methods that are actually used now. As formulated by Schauer (2010, p. 102): 'In law as in science, "compared to what?" is an important question.' Neuroscientific methods should therefore only be dismissed if it can be shown that they do not add any useful piece of evidence to the existing methods.

Even with the expected increase of direct neuroscientific evidence in 'mens rea' assessment, it will also in the future remain the task of human beings (jurors or judges) to determine 'mens rea'. In the second part of this chapter, we have summarized neuroscientific studies investigating which brain regions are involved when people try to do this. The results show that areas underlying ToM, especially the right TPJ, play a prominent role in this process, as well as the DLPFC. Compared to the descriptions from the first part of this chapter, these findings are unlikely to have any direct consequences on actual legal procedures in the forseeable future. Although certain selection criteria for jurors and judges do exist (Litteneker 1978; Lieberman & Sales, 2007), it is hard to imagine that legal decision makers will ever be selected for a given case on the basis of the activation patterns in their ToM network or DLPFC. However, these findings may indirectly contribute to the discussion by making clear that when it comes to the legal requirement of assessing 'mens rea', the involved human beings (jurors and judges) can formally also be regarded as 'measuring instruments' (with brain activity as a kind of 'display'), and, as such, they also have to be proven to be reliable, at least more reliable than any other instrument available according to the comparative view mentioned above. It is well known from psychological research that humans (including legal experts) are susceptible to a variety of cognitive biases and in many cases do not know the factors that actually determine their behaviour, although they think they know them (Nisbett & Wilson, 1977). Thus, in addition to neuroscientific findings, results from cognitive science also deserve more attention in the legal system (Busey & Loftus, 2007; Goodenough & Tucker, 2010). Two examples may illustrate this. The anchor effect is a well-known psychological effect relating to the fact that information which is mentioned first will bias people to adjust their behaviour according to that information. For example, if legal experts read cases, the number of years they suggest for the sentence is strongly influenced by the number of years suggested in the files. This is true even if they know that the number in the files has been determined by chance. Actually, the effect is still present if they themselves throw dices and write the number of years into the file case to be judged (Englich, Mussweiler, & Strack, 2006). Another, similarly frightening, example, more specifically referring to intentionality, is the phenomenon of 'choice blindness' demonstrated by Johansson and colleagues (Johansson et al., 2005). Subjects in this study were asked to choose from two photographs the face they found more attractive. Then, after a short delay, they received a photograph, purportedly the one they had chosen, with an instruction to explain the reasons why they preferred that face. In reality, however, they received the non-chosen photograph. A vast majority of subjects did not realize that they had received the non-chosen face and willingly explained which features of the (wrong) face had led them to prefer it over the other face. Even more telling with regard to real legal decisions, recently it has been shown that judges sentence much more favourably in cases immediately after they have had a food break than in cases that were treated long after a break (Danziger, Levav, & Avnaim-Pesso, 2011). This shows that even trained legal experts are not protected against irrelevant influences on their cognitive processes. The study by Schleim et al. (2011) described above further shows that even after years of specific education, legal experts use basically the same brain mechanisms as non-experts when they make legal and moral decisions. Thus, although the role of a judge normatively presupposes an objective view, judges cannot rid themselves of their human nature, and they should at least be aware of this.

In summary, we would recommend that both neuroscientists and legal experts are self-critical with regard to their own capabilities and are open to the discussion about what are the best methods for achieving the aims of legal procedures, specifically, in this context, the aim of determining whether a given defendant had a 'guilty mind' or not. Every procedure, whether neuroscientific or not, that can provably contribute to this aim should be welcome. Such an open discussion would ultimately be for the service of the optimal implementation of justice, which those citizens whose lives depend on the legal decisions rightfully demand and deserve.

The ultimate criterion for the evaluation of neuroscientific evidence should be the question of whether the legal task of 'mens rea' assessment can be improved by its use in comparison to alternative procedures that are available. We think that the answer is certainly 'yes' under some conditions, but legal experts, neuroscientists, and psychologists have to work together to specify these conditions as exactly as possible.

NOTES

1 In this chapter we will focus predominantly on *functional* neuroimaging methods because this approach is new and not yet accepted in the courts, whereas the relevance of traditional structural neuroimaging (CAT and structural MRI), that can show large brain lesions or tumours, is already acknowledged in most legal systems to a certain degree.

2 We do not consider here the ethical issues of the application of lie detection technologies, especially if their use is not voluntary, and the ensuing legal consequences regarding their actual use. The focus here is on the question of what information they can provide with regard to 'mens rea' *if* they are applied.

3 Legal systems can differ considerably between countries and even between subunits within countries (e.g. states of the USA). For specific statements on legal rules and decisions we therefore always additionally refer to their country of origin. The focus will be on the US system, which will be the most familiar system to the majority of readers.

4 This does not mean that the polygraph is not applied at all anymore. It is still widely used by private companies and by the military, especially in the USA. Polygraph evidence is also still used in certain legal contexts, e.g. in civil cases and in pre-trial hearings in the USA and other countries.

5 In Germany, the polygraph has never been used in court. However, the justification for its non-admissibility has changed. In 1954, the Federal Court of Justice ('Bundesgerichtshof') argued that the use of a polygraph would offend against human dignity (Spranger, 2009). In a new decision in 1998, however, it only argued that the validity and reliability of the polygraph is insufficient, or, to be more precise, that polygraph data have to be considered as completely ineligible evidence, at least when obtained with the usual 'comparison question test' (Seiterle, 2011).

6 Sometimes this rule is spelled 'McNaughton' rule, according to the spelling used in the original trial (The Queen *vs.* Daniel McNaughton, 1843).

7 In this chapter, we use the term 'mens rea' in a broad sense. According to Kadish (1968), this term 'is rivalled only by the term "jurisdiction" for the variety of senses in which it has been used'. He sees at least two principal categories of 'mens rea' that need to be distinguished: 'mens rea' in its special sense, referring only to the mental state which is required by the definition of the offence to accompany the act which produces the harm; and 'mens rea' in its general sense, referring to legal responsibility, which includes the typical familiar defences like insanity and infancy. The present chapter uses 'mens rea' in an even broader sense, referring to all legal situations where neuroscientific evidence could theoretically contribute to evaluate 'guilty minds' of defendants, also including procedures like the assessment of diminished culpability and lie detection.

8 In Germany, as in most other European countries, lack of control is, usually, the second criterion (apart from insight) for the insanity defence as defined in Section 20 of the German penal code: 'He acts without guilt who at the commission of the act was incapable of understanding the wrongfulness of the act or of acting in accordance with this understanding, because of mental illness, a profound, far-reaching disturbance of consciousness or because of mental retardation or because of another severe mental abnormality.' (Translation cited after Kröber, 2009.)

9 For a collection of specific cases, compare the blog 'Biosciences and the law' by Nita Farahany (http://lawandbiosciences.com/).

10 We do not want to argue from these findings that psychopathy or antisocial personality disorder should be generally regarded as an excuse in lawsuits. But knowledge about consistent brain anomalies in these populations should be considered in the same way as in other cases of mental disorders such as schizophrenia, even if it appears to be more difficult in these cases to draw the separation line between 'the sick' and 'the bad'.

11 It has been extensively debated how legal judgement is, and should be, related to moral judgement (Hart, 1958; Posner, 1998). Of course, both are not the same. Many legal prescriptions just aim at organizing societal procedures most efficiently and most smoothly, without any moral implications. But at least in the context of criminal law, it is difficult to imagine how legal judgement should be accomplished without reference to certain moral principles, whether implied in the law text itself or applied explicitly or implicitly in the considerations of jurors or judges during the trial.

12 Even more frequently, neuroscientific information is already used in civil law not considered here, e.g. in the determination of occupational disability.

13 For an overview of methodological problems of and unjustified inferences from neuroimaging results, see Walter (2009).

ACKNOWLEDGMENTS

We thank Dorit Kliemann, Sabine Müller, Kristin Prehn, Stefan Seiterle, and Andrew Wold for critical comments on an earlier version of this manuscript.

REFERENCES

American Law Institute (1962). *Model Penal Code*. Philadelphia (PA).

Aronson, J. D. (2007). Brain imaging, culpability, and the juvenile death penalty. *Psychology, Public Policy, and Law, 13,* 115–142.

Batts, S. (2009). Brain lesions and their implications in criminal responsibility. *Behav.Sci.Law, 27,* 261–272.

Beckman, M. (2004). Crime, culpability, and the adolescent brain. *Science, 305,* 596–599.

Bennett, M. (2009). Criminal law as it pertains to 'mentally incompetent defendants': a McNaughton rule in the light of cognitive neuroscience. *Australian and New Zealand Journal of Psychiatry, 43,* 289–299.

Blair, R. J. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J.Neurol Neurosurg Psychiatry, 71,* 727–731.

Brüne, M. & Brüne-Cohrs, U. (2006). Theory of mind – evolution, ontogeny, brain mechanisms and psychopathology. *Neurosci.Biobehav.Rev., 30,* 437–455.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D. et al. (2008). The neural correlates of third-party punishment. *Neuron, 60,* 930–940.

Busey, T. A. & Loftus, G. R. (2007). Cognitive science and the law. *Trends Cogn Sci., 11,* 111–117.

Carrington, S. J. & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum.Brain Mapp., 30,* 2313–2335.

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G. et al. (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia, 45,* 3105–3113.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proc.Natl.Acad.Sci.U.S.A, 108,* 6889–6892.

Daubert *vs.* Merrell Dow Pharmaceuticals (1993), 509 U.S. 579.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W. et al. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage, 28,* 663–668.

de Bruxelles, S. (2009). Man who killed wife in sleep, Brian Thomas, walks free after charges withdrawn. *The Times* (online report at http://www.timesonline.co.uk/tol/news/uk/article6925322.ece), 21 November.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making. *Pers.Soc. Psychol.Bull., 32,* 188–200.

Erk, S., Mikschl, A., Stier, S., Ciaramidaro, A., Gapp, V., Weber, B. et al. (2010). Acute and sustained effects of cognitive emotion regulation in major depression. *J.Neurosci., 30,* 15726–15734.

Feresin, E. (2009). Lighter sentence for murderer with 'bad genes'. *Nature News* (online at http://www.nature.com/news/2009/091030/full/news.2009.1050.html), 30 October.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. et al. (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition, 57,* 109–128.

Frith, C. D. & Frith, U. (1999). Interacting minds – a biological basis. *Science, 286,* 1692–1695.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38,* 11–21.

Gazzaniga, M. S. (2008). The law and neuroscience. *Neuron, 60,* 412–415.

Gewirtz, P. (1996). On 'I know it when I see it'. *Yale Law Journal, 105,* 1023–1047.

Goodenough, O. R. & Prehn, K. (2004). A neuroscientific approach to normative judgment in law and justice. *Philos.Trans.R.Soc.Lond B Biol.Sci., 359,* 1709–1726.

Goodenough, O. R. & Tucker, M. (2010). Law and cognitive neuroscience. *Annu.Rev. Law Soc.Sci., 6,* 61–92.

Grubin, D. (2010). The polygraph and forensic psychiatry. *J.Am.Acad.Psychiatry Law, 38,* 446–451.

Hare, R. D. (1991). *The Hare Psychopathy Checklist – Revised.* Toronto, Ontario: Multi-Health Systems.

Hart, H. L. A. (1958). Positivism and the separation of law and morals. *Harvard Law Review, 71,* 593–629.

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Curr.Biol., 17,* 323–328.

Hirono, N., Mega, M. S., Dinov, I. D., Mishkin, F., & Cummings, J. L. (2000). Left frontotemporal hypoperfusion is associated with aggression in patients with dementia. *Arch.Neurol., 57,* 861–866.

Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310,* 116–119.

Juhasz, C., Behen, M. E., Muzik, O., Chugani, D. C., & Chugani, H. T. (2001). Bilateral medial prefrontal and temporal neocortical hypometabolism in children with epilepsy and aggression. *Epilepsia, 42,* 991–1001.

Kadish, S.H. (1968). The decline of innocence. *Cambridge Law Journal, 26(2),* 273–290.

Killen, M., Lynn, M. K., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: morally-relevant theory of mind. *Cognition, 119,* 197–215.

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia, 46,* 2949–2957.

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

Kröber, H.-L. (2009). Concepts of intentional control. *Behav.Sci.Law, 27,* 209–217.

Langleben, D. D., Loughead, J. W., Bilker, W. B. et al. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Hum.Brain Mapp., 26,* 262–272.

Lieberman, J. D. & Sales, B. D. (2007). *Scientific Jury selection.* Washington, DC: American Psychological Association.

Litteneker, R. J. (1978). Disqualification of federal judges for bias or prejudice. *University of Chicago Law Review, 46(1),* 236–268.

Loftus, E. F., Doyle, J. M., & Dysert, J. (2008). *Eyewitnes Testimony: Civil & Criminal.* Charlottesville, VA: Lexis Law Publishing.

Miller, G. (2010). fMRI lie detection fails a legal test. *Science, 328,* 1336–1337.

Mohamed, F. B., Faro, S. H., Gordon, N. J. et al. (2006) Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation – initial experience. *Radiology, 238,* 679–688.

Moll, J. & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends Cogn Sci., 11,* 319–321.

Moriarty, J. C. (2008). Flickering admissibility: neuroimaging evidence in the U.S. courts. *Behav.Sci.Law, 26,* 29–49.

Müller, S. & Walter, H. (2011). Machen die Neurowissenschaften einen Paradigmenwechsel im Strafrecht erforderlich? In G. Scharifi (ed.) *Brauchen wir eine neue Ethik? Herausforderungen der Moral durch die Neurowissenschaft* (pp. 143–164) [Do the neurosciences make a paradigm shift in criminal law necessary? In G. Scharifi (ed.) *Do We Need New Ethics? Neuroscientific Challenges for Morality*]. Paderborn (Germany): Mentis.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Piaget, J. (1932). *The Moral Judgment of the Child.* New York: Free Press.

Posner, R. A. (1998). The problematics of moral and legal theory. *Harvard Law Review, 111,* 1637–1717.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci., 1,* 515–526.

Raine, A. & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Soc.Cogn Affect.Neurosci., 1,* 203–213.

Roper *vs.* Simmons (2005), 543, U.S. 551.

Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends Cogn Sci., 10,* 419–423.

Saxe, R. (2006). Uniquely human social cognition. *Curr.Opin.Neurobiol., 16,* 235–239.

Saxe, R. & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *Neuroimage, 19,* 1835–1842.

Schardt, D. M., Erk, S., Nusser, C., Nothen, M. M., Cichon, S., Rietschel, M. et al. (2010). Volition diminishes genetically mediated amygdala hyperreactivity. *Neuroimage, 53,* 943–951.

Schauer, F. (2010). Neuroscience, lie-detection, and the law: contrary to the prevailing view, the suitability of brain-based lie-detection for courtroom or forensic use should be determined according to legal and not scientific standards. *Trends Cogn Sci., 14,* 101–103.

Schleim, S., Spranger, T. M., Erk, S., & Walter, H. (2011). From moral to legal judgment: the influence of normative context in lawyers and other academics. *Soc.Cogn Affect.Neurosci., 6,* 48–57.

Schleim, S., Spranger, T. M., & Walter, H. (2009). *Von der Neuroethik zum Neurorecht?* Göttingen (Germany): Vandenhoeck & Ruprecht. [*From Neuroethics to Neurolaw?*].

Schopp, R. F. (1991). *Automatism, Insanity, and the Psychology of Criminal Responsibility.* Cambridge: Cambridge University Press.

Segrave, K. (2004). *Lie Detectors: A Social History.* Jefferson, NC: McFarland & Company.

Seiterle, S. (2010). Hirnbild und 'Lügendetektion' [Brain image and 'lie detection'.] Berlin (Germany): Duncker & Humblot.

Seiterle, S. (2011). Zur Eignung von fMRT-basierten 'Lügendetektoren' im Strafprozess [On the suitability of fMRI-based 'lie detection' in criminal proceedings]. *Nervenheilkunde, 30,* 437–442.

Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends Cogn Sci., 12,* 48–53.

Spranger, T. M. (2009). Legal implications in connection with the generation and usage of neuro-scientific findings. *Journal of International Biotechnology Law, 6,* 228–234.

The Queen *vs.* Daniel McNaughton (1843), 4 St. TR. (N.S.) 847.

U.S. *vs.* Scheffer (1998), 523 U.S. 303.

U.S. *vs.* Semrau (2010), No. 07-10074 MI/P.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P. et al. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage, 14,* 170–181.

Wagner, U., N'Diaye, K., Ethofer, T., & Vuilleumier, P. (2011). Guilt-specific processing in the prefrontal cortex. *Cereb.Cortex, 11,* 2461–2470.

Walter, H. (2001). *Neurophilosophy of Free Will: From Libertarian Illusion to the Concept of Natural Autonomy.* Cambridge, MA: MIT Press.

Walter, H. (2009). Was können wir messen? Neuroimaging – eine Einführung in methodische Grundlagen, häufige Fehlschlüsse und ihre mögliche Bedeutung für Strafrecht und Menschenbild. In S. Schleim, T. M. Spranger, & H. Walter (eds.) *Von der Neuroethik zum Neurorecht?* (pp. 67–103) [What can we measure? Neuroimaging – an introduction into methodological bases, frequent false conclusions, and its possible relevance to criminal law and the idea of man. In S. Schleim, T. M.

Spranger, & H. Walter (eds.) *From Neuroethics to Neurolaw?*]. Göttingen (Germany): Vandenhoeck & Ruprecht.

Walter, H. (2011). Contributions of neuroscience to the free will debate: from random movement to intelligible action. In R. Kane (ed.) *Oxford Handbook of Free Will* (pp. 515–529). Oxford: Oxford University Press.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J.Cogn Neurosci., 16,* 1854–1863.

Walter, H., Kalckreuth, A., Schardt, D. M., Stephan, A., Goschke, T., & Erk, S. (2009). The temporal dynamics of voluntary emotion regulation: immediate and delayed neural aftereffects. *PLoS ONE, 4,* e6726.

Weber, S., Habel, U., Amunts, K., & Schneider, F. (2008). Structural brain abnormalities in psychopaths – a review. *Behav.Sci.Law, 26,* 7–28.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev., 72,* 655–684.

Wettstein, R. M., Mulvey, E. P., & Rogers, R. (1991). A prospective comparison of four insanity defense standards. *Am.J.Psychiatry, 148,* 21–27.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.

Yang, Y. & Raine, A. (2009). Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. *Psychiatry Res., 174,* 81–88.

Young, L. Camprodon, J., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc.Natl.Acad.Sci.U.S.A, 107,* 6753–6758.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proc.Natl.Acad.Sci.U.S.A, 104,* 8235–8240.

Young, L. & Saxe, R. (2009). Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47,* 2065–2072.