

# Neuroimaging and Responsibility Assessments

Nicole A Vincent

Received: 19 July 2008 / Accepted: 18 December 2008 / Published online: 21 January 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** Could neuroimaging evidence help us to assess the degree of a person's responsibility for a crime which we know that they committed? This essay defends an affirmative answer to this question. A range of standard objections to this high-tech approach to assessing people's responsibility is considered and then set aside, but I also bring to light and then reject a novel objection—an objection which is only encountered when functional (rather than structural) neuroimaging is used to assess people's responsibility.

**Keywords** Moral responsibility · Legal responsibility · Capacity-theoretic conception of responsibility · Capacitarian theory of responsibility · Mental capacity · Capacity responsibility · Neuroimaging · fMRI · Modal fallacy · Automatic functions · Theory to the best explanation · Roper v. Simmons [2005]

## Christopher Simmons Murders Shirley Crook<sup>1</sup>

In the early hours of the morning on September 9th, 1993 in St. Louis County, Missouri, a 17 year old Christopher Simmons and his younger friend Charles Benjamin broke into the home of Mrs Shirley Crook who was asleep at the time. Simmons and Benjamin burgled her home and they bound the now awakened woman's arms and legs with electric cable and leather straps, covered her face with a towel and duct tape, loaded her into the back of her own mini-van, and drove to the nearby Castlewood State park where Simmons threw her off a railway bridge and into the Maramec River down below. Hours later Mrs. Crook's lifeless body was retrieved from the river by fishermen, and the medical examiner determined that drowning was the cause of her death.

That Christopher Simmons killed Shirley Crook is not in dispute. His role in the crime was clearly established in the investigations that followed, and these investigations also revealed both the pre-meditated nature of the crime (witnesses testified that Simmons had talked about wanting to burgle and murder someone for some time) and the fact that

---

N. A. Vincent (✉)  
Philosophy, TBM, TU Delft,  
P.O. Box 5015, 2600 GA Delft, The Netherlands  
e-mail: n.a.vincent@tudelft.nl

---

<sup>1</sup> This case summary is compiled from several sources, and I provide this single citation to avoid many repetitive citations. The sources used were [1–4].

immediately afterwards he felt not remorse or guilt but pride and a sense of accomplishment (at school later that day he bragged to his friends about the murder which he had just committed).

### Age, Mental Capacities and Three Problems Associated with Individualized Behavioural Assessments of Responsibility

However, although there is no doubt that Christopher Simmons killed Shirley Crook, we may still legitimately ask whether he is fully responsible for what he did. The primary consideration which mitigates against attributing full responsibility to him for this crime<sup>2</sup> [5] is age-related—on at least some definitions Simmons was still a minor when he committed the crime, and since we normally suppose that minors are less than fully responsible for what they do on account of their immaturity, we may therefore also be tempted by the suggestion that Simmons too was not fully responsible for his crime. However, several matters complicate the issue, and perhaps the most germane is the fact that when he committed the crime Simmons was an adolescent approaching adulthood, which perhaps suggests an important disanalogy between him and other minors that counts against treating him as we would a younger child. Admittedly though, age is a very rough guide to how mature a person really is—some people mature a lot faster than others—and so if responsibility is indeed a function of maturity then perhaps the law should *individually* assess people's responsibility—when so much is at stake (how severely Simmons will be punished) we should surely aim for greater precision and accuracy!

One way to individually assess people's responsibility might look like this. Working under the

<sup>2</sup> When I talk of “*attributing* responsibility to him for this crime” my concern is with what might be called his *outcome responsibility* (i.e. what states of affairs he brought about; e.g. [5]:555), not with what Christopher Kutz and H.L.A. Hart refer to as his *liability responsibility* (i.e. what we should do to him-, how we should treat him-, how we should hold him responsible-, or even how he should take responsibility for his crime; [6]:211; [7]:549). To avoid equivocation, it is crucial to keep a clear distinction between the many different senses of the term “responsibility”, or what I elsewhere refer to as the many different “responsibility concepts” ([8]:115), more of which are briefly mentioned in the following footnote.

assumption that a person's responsibility is a function of such things as their *cognitive* and *volitional* capacities—i.e. respectively, their capacity for making sound moral judgments and their capacity for exercising appropriate self-control<sup>3</sup>—we could attempt to gauge how responsible a person is by using the following *behavioural* approach; for instance, we could set them moral judgment tasks and self-control challenges, watch how well they perform, and presumably the better they perform at these tests the greater will be their cognitive and volitional capacities. Alternatively, we could talk to others who know them well—for instance, their family, friends, neighbours, teachers and peers—in an effort to build up an accurate picture of their cognitive and volitional capacities from these people's biographical accounts.

By using this behavioural approach we would be trying to determine not just whether the person under assessment *can make* moral judgments and whether they *can control* themselves, but rather whether they *can make sound* moral judgments and whether they *can appropriately* control themselves. After all, what matters for responsibility is not just that one can make *some* moral judgments, but rather that one's moral judgments are *sound*, and as Susan Wolf points out, for this to happen an agent must be “able to ‘track’ the True and Good in her value judgments” ([12]:211; also see [11]). One way to conceive of the point that Wolf is making here is that in addition to the *formal* requirements upon how one's reason must operate in order to be valid, there are also *substantive* requirements on the content of the premises from which one reasons or which form the background assumptions

<sup>3</sup> I suggest these particular capacities because of what seems to me like an obvious link between maturity and responsibility—presumably, this is at least part of the reason why maturity is meant to be relevant to responsibility—but also because of the prominence of these two mental conditions (or other closely related conditions) in the relevant literature (e.g. [9–16]). However, as I argue elsewhere—and I believe that others would agree on this point (e.g. see [17]:71; [18]:38–9)—responsibility does not *just* hinge on facts about a person's mental capacities (their *capacity responsibility*), since in the least it probably also depends on such things as how they acted and whether there was a connection between their action and the outcome for which they are allegedly responsible (their *causal responsibility*), whether in acting as they did they violated their *role responsibilities*, and many other things such as whether they are responsible for the fact that they now lack those mental capacities ([19]:202–3).

that inform one's moral judgment—or what Neil Levy refers to as “moral knowledge” [20]—since these substantive issues also bear on whether one's otherwise flawless reasoning processes will be sufficient to bestow upon one the status of a responsible moral agent. Similarly, what matters for responsibility is not just that one can control oneself in *some* way, but rather that one can control oneself in *the right* way—one way of putting this point is that one's actions must issue from a mechanism that is moderately responsive to reasons [9].<sup>4</sup> Thus, what such a behavioural approach would attempt to do is to determine whether the person being tested can make *sound* moral judgments (e.g. whether they are able to give the right sorts of answers to moral questions and whether one can justify those answers appropriately) and whether they can *appropriately* control themselves (e.g. whether they are able to resist temptations to do what they think should be resisted).

However, I see at least three problems with using this behavioural approach to the individual assessment of people's responsibility. As regards the first suggestion (i.e. administering tests), given that the outcome of such a test may eventually bear on how severely the person is punished, subjects would have every reason to pretend that their capacities are a lot lower than what they really are—i.e. they would have every reason to intentionally flunk such tests by giving the most ridiculous answers to the moral judgment tasks and by yielding to even the weakest of temptations, to make it seem as if their cognitive and volitional capacities are severely retarded. As regards the second suggestion (i.e. talking to people who know them), given that their family and friends may want to defend them and that their enemies may want to see them perish, if what we wanted was an accurate, precise and impartial method of gauging a person's responsibility then this would not be the best alternative since there is simply too much opportunity for everyone to lie about what the person whose responsibility is being assessed is really like.

Most importantly though—and this point counts against both of these suggestions on purely philosophical grounds—strictly speaking such tests are actually incapable of revealing anything about what capacities a person lacks (about what they *cannot* do). The mere

fact that someone has always acted badly—irrespective of whether this is something that we ascertain from repeated *tests* or from the *stories* that others tell about them, and irrespective of whether this relates to their *moral judgment* or to their *self-control*—may simply only show that they are *lazy* rather than that they are *incapacitated*—i.e. that they *do not* make the right moral judgments rather than that they *can not* make the right moral judgments, or that they *do not* control themselves rather than that they *can not* control themselves. For instance, Bernadette McSherry also points out that “it is impossible to devise an objectively verifiable test to determine when an accused *could not* control him or herself and when he or she merely *would not*” ([25]:188, emphasis added), and Robert Sapolsky highlights this same point by drawing attention to the difficulties associated with “distinguishing between an *irresistible* impulse and one that is to any extent resistible but which *was not resisted*” ([13]:1790, emphasis added).

My point is that even if the stories told by others revealed that Christopher Simmons had always misbehaved ever since he was little, that would not necessarily have shown that he *couldn't* distinguish right from wrong or that he *couldn't* help acting that way—as we all know, children often knowingly misbehave, and hence such stories are perfectly compatible with the possibility that Simmons possessed those capacities all along but that he simply *didn't* use them. Alternatively, even if repeated behavioural tests show that Christopher Simmons consistently makes the wrong moral judgments and that he consistently yields to even the weakest of temptations, this will not necessarily show that he *can't* distinguish right from wrong or that he *can't* control himself, since this behaviour can also be explained by the alternative hypothesis that he simply *doesn't* use capacities which for all we know he does actually possess; perhaps the reason why he acts impulsively and why he doesn't resist temptation is because he is lazy (i.e. maybe he simply doesn't try hard enough).

This third problem is an instance of the so-called *modal fallacy* in which certain types of conclusions about what is *possible* are invalidly derived from premises about what is *actual*—in this case, observations about what doesn't occur (about a person's *behaviour*) are erroneously taken to entail conclusions about what can't occur (about their *incapacity*). Some

<sup>4</sup> Alternatively, see [21] or one of [22–24] for concise summaries of Fischer and Ravizza's position.

conclusions about what is *possible* can be validly derived purely from premises about what is *actual*—for instance, the fact that it *is* raining does entail that it is *possible* that it is raining. However, in the case at hand what we have is something more like this: the fact that it *is not* raining entails that it is *not possible* that it is raining; and this latter sort of derivation of a conclusion about what is possible from a claim about what is actual is patently invalid. Thus, quite apart from the fact that the person being tested and that others who tell us about them can deceive and lie, this modal fallacy poses a serious philosophical problem for using such behavioural approaches in the individual assessment of people's responsibility.

Hence, for at least these three reasons—i.e. (1) because people can pretend that their cognitive and volitional capacities are impaired when in fact they are fully intact, (2) because other people's biographical accounts of what someone else is like may be inaccurate, and (3) because of the modal fallacy that was just described—I find such behavioural assessments of responsibility wanting.

### The Promise of Neuroimaging

However, neuroimaging may seem to offer us a better solution—a high-tech way of individually assessing people's responsibility that does not run into these three problems—by making it possible to directly inspect people's brains and to bypass the filter of their own and of other people's agendas. Speaking broadly and schematically, and working yet again under the earlier assumption that a person's responsibility is determined in part and amongst other things by what responsibility-relevant capacities they possess, neuroimaging could conceivably help by first revealing precisely which brain mechanisms are required for responsible moral agency, and then by helping us to check whether a specific individual whose responsibility is in need of individual assessment possesses those brain mechanisms. The basic idea here is simple: to be a responsible moral agent one must have the right mental capacities, but since mental capacities are implemented in brain mechanisms (in brain “hardware”), to be a responsible moral agent one must have the right brain mechanisms, and *that*—i.e. whether the person whose responsibility is being assessed has those brain mechanisms—is precisely

what neuroimaging would be used for in this high-tech approach to the individualized assessment of responsibility.<sup>5</sup>

Something like this is already being done in the moral cognition literature, where researchers first try to reveal which parts of the brain are used in various kinds of moral judgment tasks, and this information can then be used in attempts to shed light on why specific individuals make queer moral judgments (e.g. see [26]; or the collection of articles introduced by [27]). Similarly, it is conceivable that neuroimaging could also be used to study the neural basis of moral responsibility to determine what “common core of mental capacities [a] person must have in order to be a responsible or competent moral agent” ([19]:202, emphasis removed) and then to use this knowledge to help us individually assess specific people's responsibility by checking whether they have those capacities or not, except here our interests would be somewhat broader than they are in the area of moral cognition if only because moral responsibility hinges on volitional as well as cognitive capacities. Although at present it is not clear *precisely* which mental capacities are required for responsible moral agency—for instance, whether a capacity for rationality is all that is required (e.g. [18]), or whether one must also have the right emotional/affective capacities (e.g. [28]), or maybe even some more finely-grained capacity that will be revealed in the future by neuroimaging studies (e.g. [19]:203; [29])—presumably, with the help of neuroimaging in the future we will discover precisely which capacities play an indispensable role in competent moral agency.

That neuroimaging could be used to help us individually assess people's responsibility in roughly this manner is not a novel suggestion. For instance, Walter Glannon writes that “functional magnetic resonance imaging (fMRI) can display the structure and function of the brain regions that regulate our capacity for impulse control” ([17]:68), and he suggests that an “fMRI image showing significantly reduced metabolic activity in ... this part of the brain, could excuse [a person] from [a] charge of responsibility [o]n the basis ... that she lacked [a relevant]

<sup>5</sup> I am working here under the assumption that our mental capacities are implemented within our brains, though this assumption might perhaps be challenged by (1) dualists, (2) those who claim that other parts of the nervous system also play an important role, and by (3) proponents of the extended mind hypothesis.

capacity” ([17]:75). Similarly, Adrian Raine and Yaling Yang summarize a growing body of empirical data which shows significant structural and functional differences between the brains of normals when compared to the brains of “delinquent, criminal, violent and psychopathic populations”, and towards the end of their paper they suggest that although “[p]sychopaths may not be...insane...if they lack the capacity for the feeling of what is moral due to neurobiological impairments beyond their control, [then they might not be] fully responsible for their criminal behavior” ([30]:211, emphasis removed).

The advantages of the neuroimaging approach to the individualized assessment of people’s responsibility over the behavioural approaches outlined earlier can be summarized briefly. Firstly, within this approach there is no room for deceit and lies to spoil the validity of our findings, because our findings do not rely on anyone’s veracity—i.e. this approach can not be fooled by someone who intentionally flunks their moral judgment and self-control tests, and it does not rely for its validity upon the accuracy of other’s biographical accounts of the person whose responsibility is in need of being individually assessed. Secondly, to the extent that the mechanisms present within our brains—i.e. the “hardware” of our brains—determine our mental capacities,<sup>6</sup> and further, to the extent that neuroimaging provides us with a direct view of our brains,<sup>7</sup> the neuroimaging approach enables us to directly view what mental capacities a person really has, and so it is also immune to the modal fallacy problem which plagues the described behavioural approaches.<sup>8</sup>

### Some Problems Associated with the Neuroimaging Approach

Naturally, nobody is implying that this promising high-tech solution to the individualized assessment of

<sup>6</sup> I comment on some of the limitations behind this assumption in note 5 above.

<sup>7</sup> I discuss this point in detail below—see note 9 and the surrounding text as well as the section below entitled “Brief Replies to the Problems Identified in the Previous Section”.

<sup>8</sup> I discuss a complication with this last point—a complication that applies to the use of *functional* neuroimaging technology in this context—in the section below entitled “The Modal Fallacy and Functional Neuroimaging”.

people’s responsibility is a “silver bullet” which will answer all of our pertinent questions, or that it would be completely free of problems.

For instance, Glannon mentions a long list of caveats including: (1) not all people with brain abnormalities commit crimes, and so we ought not conclude in any given case that the brain abnormality which is observed in a particular individual is what caused them to commit their crime ([17]:77); (2) no scan of the brain performed after the event can establish with any degree of certainty what state that person’s brain was in at the time when they committed the crime, and this temporal aspect may also limit our ability to determine whether it is legitimate to attribute responsibility to someone for a crime which we know that they committed ([17]:80); (3) fMRI images and images from other current functional neuroimaging technologies are not direct images of brain activity, but theory-laden representations of the outcomes of statistical analyses performed upon data about metabolic activity in the brain gathered in highly controlled settings when subjects respond to often very artificial and carefully hand-crafted questions ([17]:80–1);<sup>9</sup> and (4) that responsibility is a normative notion “reflecting social conventions and expectations about how people can or should act” and not just empirical findings about a

<sup>9</sup> An fMRI scan is by no means a *direct* image of neural activity since it only reveals the blood oxygenation level changes in the brain, and so inferences must be made from observations about blood oxygenation changes to claims about which parts of the brain are being used, but yet a great many assumptions need to be granted in order for these inferences to be accepted. Furthermore, fMRI images reveal areas of the brain which are *associated* with the performance of that task; but although the *association* might be that those parts of the brain *comprise/implement* (partly or fully) the respective function or that in some other way they *support* that function, it might also be that those parts of the brain activate as a *consequence* (i.e. as effects rather than as causes or constituents) of the performance of that function. These same points and many others—i.e. all cautioning us about the fact that modern neuroimaging techniques are still in their infancy, that they are still an inexact science, and hence that we must exercise a healthy degree of scepticism or at least caution about what conclusions we draw from the data gathered using these techniques—are also made elsewhere (for instance, see [31–34]).



person's mental capacities, and hence we should not expect neuroimaging to provide answers to all of the relevant questions which must be answered in order to ascertain a person's responsibility ([17]:80–1). On the other hand, Raine and Yang point out that our “understanding of the neural basis to moral decision-making is clearly still in its infancy”, and that “social risk factors” must surely also play a role in disposing individuals to certain types of behaviour ([30]:211)—i.e. they too are reserved about just how much neuroimaging can revolutionize individualized assessments of responsibility.

Similar comments and related concerns about the extent to which neuroimaging can revolutionize individualized assessments of responsibility are also voiced by others, or else they are implied by what they have to say. For instance, like Glannon, Stephen Morse also points out that responsibility attributions depend at least as much on “normative moral, political, social and ultimately legal constitutional question[s] about which behavioral science and neuroscience must...fall silent” ([18]:48), as they do on empirical facts about people's brains. For instance, he argues that only the law may say precisely what threshold of capacity is sufficient for a person to pass the “minimal rationality” test and thus to be deemed a responsible agent; and relatedly, that only it may determine how much compulsion is too much for anyone to bear, and hence that it is a matter for those working in the legal profession and not for neuroscientists to decide how much compulsion must be present before a person's responsibility becomes compromised ([18]:38). Furthermore, as regards meting out punishment to those who have already been deemed responsible for some state of affairs, this too is at least partially a normative issue because only the law is in a position to determine what is and what is not an *appropriate* level of punishment for a given criminal violation—for instance, whether the appropriate retributive punishment for a given offence should be some harsh words, a slap on the wrist, a fine, imprisonment, ten lashes of the whip, twenty lashes, or maybe even execution. On Morse's account, the law's responsibility practices are about norm-setting and not just about fact-finding, and this is an important reason why neuroimaging can't answer all of the relevant questions that must be answered in order to determine the extent of some-

one's responsibility for something that they have done.<sup>10</sup>

The list of doubts and problems that limit the extent to which modern neuroimaging techniques can revolutionize how we individually assess people's responsibility does not end here. Firstly, given neural plasticity—i.e. the fact that the same function might be implemented in different people's brains in radically different ways<sup>11</sup>—we should not conclude too hastily that someone lacks particular capacities just because they lack the related brain mechanism(s) since those capacities may be implemented in *their* brain in a very different way to what we expect.

Secondly, as I point out elsewhere ([19]:202), although brain scans may reveal that someone lacks a capacity which is required for responsible moral agency, if we have reason to believe that they are responsible for that incapacity<sup>12</sup> then the mere fact that they now have this incapacity will not necessarily exculpate them. For instance, a drunk driver who causes a car accident can not excuse themselves by citing the fact that they lacked the capacity to drive safely and hence that it was unreasonable for us to

<sup>10</sup> In general Morse seems to be very skeptical about the relevance of neuroscience to the law's responsibility practices. In fact, at one point he even “contends that neuroscience is largely *irrelevant*” to how the law arrives at its decisions concerning “responsibility ascriptions and just punishment” ([18]:34, emphasis added). I take it that Morse's point here is not just *descriptive* but also *normative*—i.e. that he is not merely *describing* what the law currently does (e.g. that neuroimaging still plays a relatively minor role within the law's responsibility practices), but that on his account the law *need not* alter its practices in light of what the cognitive neurosciences are revealing about the human mind. I will not address Morse's general scepticism about the relevance of neuroscience to the law's responsibility practices here, other than to point out that many of the things which he says show not that neuroscience is irrelevant, but rather that it simply can't answer all of the questions.

<sup>11</sup> Glannon cites the case of a female patient who, while possessing only a quarter of the cerebral cortex, still never the less possessed “no significant cognitive deficit” ([17]:79); what we have here is someone whose brain was significantly different to other people's brains—missing three quarters of the cerebral cortex is a *significant* difference on *any* account—but yet she did not lack the usual mental capacities.

<sup>12</sup> In what follows I mention just one way in which someone might come to be responsible for their incapacity—namely, by doing something that temporarily or permanently diminishes their capacities (e.g. getting drunk)—however there may be other ways too—for instance, someone might be responsible for their incapacity if they negligently fail to develop it.

expect them to drive safely in the first place in light of this incapacity. Rather, since they are responsible for their own incapacity, we would normally think that they are therefore not entitled to cite that incapacity as an exculpatory factor, and in fact we may even think worse of them for doing something that jeopardized their capacity to be a responsible agent. Thus, if someone is responsible for their current incapacity, then the mere fact that they now have this incapacity will not necessarily exculpate them, since usually only incapacities for which we are not responsible can play this exculpatory role.

Thirdly, and relatedly, although a neuroimaging scan may reveal that a person lacks some responsibility-relevant mental capacity, the mere fact that they have this incapacity will not necessarily exculpate them if we have good reason to suppose that they should have avoided putting themselves in a situation where their incapacity would pose a problem. For instance, although a driver that runs over a pedestrian as a consequence of the sudden onset of an epileptic fit may have had little if any control over their car once their epileptic fit came on, they may still never the less be responsible for the injuries and for damage that they cause if we have good reason to suppose that they ought not to have sat behind the steering wheel of their car in the first place ([32]:316; [35]).<sup>13</sup> Whether it *is* or whether it *is not* reasonable for such people to drive cars depends on a great many things,<sup>14</sup> however this should not distract us at present because the point that is currently being made is only that the mere fact that someone lacks a given capacity need not necessarily exculpate that person, especially when they knew about their incapacity and when they could reasonably have been expected to take the necessary measures to prevent that incapacity from becoming a problem.

<sup>13</sup> Stephen Morse runs a similar argument when he asks us to consider a person who is prone to explode in uncontrollable and violent fits of rage whenever they think that someone else is looking at them the wrong way—of such a person Morse says the following: “If you know you’re like that, maybe you better not go to bars” [36].

<sup>14</sup> It is conceivable that in *some* situations a person may be morally justified in exposing other road users to the risk that they will suffer a sudden epileptic fit and cause a road accident to occur—for instance, perhaps this might be the case if the risk of the onset of an epileptic fit is sufficiently small, if the cost of not driving on this occasion is too large (e.g. they need to drive their mortally sick child to hospital), and if no better options are available.

A final reason why we may not be prepared to excuse a person for something that they have done, even though they only did it because of their lack of some important responsibility-relevant mental capacity, is because their lack of that capacity may be seen as evidence for their badness and thus for their even greater culpability, rather than as evidence for their madness and thus for their reduced responsibility. For instance, Marga Reimer [37] points out that evidence from neuroimaging studies which reveals the underpinnings of psychopathy—for instance, James Blair’s [38, 39] data which show that psychopaths’ amygdala function very differently to the amygdala of non-psychopaths—is compatible with two radically different interpretations each of which has completely opposite ramifications for our assessment of psychopaths’ responsibility. On one interpretation, the fact that psychopaths’ amygdala function differently to normal people’s amygdala (and hence that perhaps on account of this different functional profile they lack the capacity to truly realize that hurting people is wrong) is evidence for their *madness*—i.e. that they suffer from a kind of mental deficit, disorder or illness which exculpates them of their responsibility. However, on another interpretation such scans provide conclusive evidence for the psychopath’s *badness*—for instance, we might imagine someone saying something like this: “Just look at those scans! See? It’s no wonder that psychopaths act like that. Their brains are wired-up to not respond to other people’s distress. At the most basic level they are wired-up to be wicked individuals who do not care about others.” Now, Reimer’s point is not that the latter interpretation is preferable to the former, but simply that such data is *neutral* between these two interpretations (e.g. [37]:188, 191–2). However, this seems to be precisely the crux of Heidi Maibom’s objection to the use of neuroimaging evidence in the service of exonerating psychopaths (and presumably others who have similarly reduced responsibility-relevant capacities) of their responsibility when she writes that “we cannot...simply excuse people for being bad” ([40]:168). Maibom’s point here seems to be that the fact that someone lacks a capacity which is required for responsible moral agency is not a reason to exculpate them, but quite to the contrary it is rather a reason to damn them even harder, since with such neuroimaging data at hand we now have conclusive evidence that they are bad people.

### Brief Replies to the Problems Identified in the Previous Section

The previous section listed ten worries and problems with the suggestion that neuroimaging can help us to individually assess people's responsibility: (1) that not all people with abnormal brains commit crimes; (2) that our understanding of how the human brain works is still very rudimentary; (3) that brain plasticity might make it difficult to diagnose who has which capacities; (4) that methodological and technological problems with current neuroimaging techniques cast doubt over the usefulness of neuroimaging data; (5) that we can't go back in time and check what capacities a person had at the time when they committed their crime; (6) that neuroimaging evidence for a person's incapacity may actually damn them even harder rather than exculpating them; (7) that social factors, and not just neurological impairments, also play some role in determining our behaviour; (8) that responsibility assessments also depend in part on normative assumptions which are at least partially independent of what cognitive neuroscience tells us about the human mind; (9) that people might be responsible for their own incapacity and thus that they might be responsible for what they do on account of that; and (10) that people who know about their own incapacities may be responsible for what they do if they fail to avoid situations in which those incapacities may become a problem.

This is not a trivial list of problems. However, at the same time I do not think that any of them are show-stoppers either, since in time some of these problems will be overcome through scientific and technological advancements, and others apply equally to the behavioural approach which was discussed above and hence such problems can not provide us with a reason to *prefer* the behavioural approach over the neuroimaging approach—i.e. we will not be any further behind if we adopt the neuroimaging approach than if we only used the behavioural approach.<sup>15</sup> For

<sup>15</sup> Perhaps if the neuroimaging approach was more prone to suffering from these problems than the behavioural approach, or if these problems could be dealt with more easily in the context of the behavioural approach than in the context of the neuroimaging approach, then citing such problems could provide reasons to prefer the behavioural approach over the neuroimaging approach, however I see no reason to think that this is so.

instance, the first three items will hopefully disappear off this list once we gain a better understanding of how the human brain works, and in any case given the brain-based causes of human behaviour, there is good reason to suppose that in the long term among the best scientific explanations for human behaviour will indeed be those that are informed by research from the field of cognitive neuroscience (e.g. [41]:61). As regards the fourth item, although current neuroimaging techniques are still in their infancy, given the track record of science and technology—i.e. their steady advancement—we can probably safely assume that better techniques will eventually come along which will not suffer from these deficiencies. The fifth item afflicts the neuroimaging approach to no greater an extent than it afflicts the behavioural approach—i.e. those who wish to assess a person's responsibility through behavioural tests also can't go back in time to check what state that person's brain was in at the precise moment at which they committed the crime—and so this is not a specific problem with the neuroimaging approach. And in addition to the fact that the last four items on this list also apply equally to the behavioural approach as they do to the neuroimaging approach, these items are merely a more detailed restatement of a claim which I have repeatedly made throughout this essay and which I do not contest—namely, that responsibility does not *just* hinge on mental capacities; however, contrary to what Morse has to say on this matter (see note 10 above), this does not entail that neuroimaging is irrelevant to responsibility assessments, but only that neuroimaging evidence has limited utility (i.e. it only solves part of the puzzle) since other considerations also impact on our assessments of people's responsibility.

This only leaves the sixth item on the list—i.e. Maibom's claim that neuroimaging data which shows that someone lacks a responsibility-relevant capacity actually damns that person even further rather than exculpating them, since on her account such data conclusively shows that this person is bad! However, strictly speaking this is not really a critique of the neuroscientific approach *per se*, but it is rather an instance of a broader concern with *all* capacitarian theories of responsibility. Any approach which claims to have found a way of revealing people's mental capacities (irrespective of whether it claims to do this by using behavioural tests or by using neuroimaging data), and that by doing this it can help us establish



the degree of their responsibility, will have to answer Maibom's objection because that approach will have taken it for granted that responsibility is in part a matter of what capacities a person has. But since Maibom's objection targets *all* theories which hold that a person's responsibility depends in part upon their mental capacities, and not just my own contention that neuroimaging could be used to help us assess a person's responsibility by helping to reveal their mental capacities, I will therefore set aside the task of defending the broader capacitarian approach.<sup>16</sup>

This section's discussion was intended to quickly ward off some common doubts about whether neuroimaging can help us to individually assess people's responsibility. As I said at the beginning of the previous section, nobody is implying that this promising high-tech solution to the individualized assessment of people's responsibility is a "silver bullet" which will answer all of our pertinent questions, or that it is completely free of problems. However, the problems which afflict it are (or at least they *promise* to be in the medium to longer term) no greater than the problems which afflict the behavioural approach, and yet it manages to overcome some serious problems which afflict the behavioural approach and for this reason I do not think that we should give up yet on the idea of using neuroimaging to help us assess people's responsibility.

### The Modal Fallacy and *Functional* Neuroimaging

Up to this point I have painted a generally positive picture of how neuroimaging could improve our ability to accurately assess people's responsibility, but there is still one further worry that needs to be addressed. It is a worry that applies specifically to *functional* (as opposed to structural) neuroimaging, and it will remain relevant even in the future once our

<sup>16</sup> In fact the capacitarian approach is defended by others. For instance, the views of such authors as John Martin Fischer & Mark Ravizza, Daniel Dennett, Susan Wolf and Patricia Smith-Churchland (all cited earlier in this essay) are all capacitarian in the right sense since each (using their own terminology) sees responsibility as a matter of whether the agent has the capacity to guide their actions in the right manner (e.g. in light of right reasons) rather than for instance as a matter of the metaphysical freedom from being determined by causality (e.g. [42]). On their accounts the fact that someone lacks this sort of capacity can exculpate them rather than showing that they are bad.

science and technology have advanced beyond their present point, and so in this section I will focus first on describing this worry, and then on showing how it too can be dealt with.

### A Question

Let me begin with the following question: could a *functional* scan of a person's brain help us assess that person's responsibility? For instance, could a set of images which conclusively show significantly reduced function in parts of the brain that undeniably play an indispensable role in underwriting the capacity for *sound moral judgment* perform this exonerating function?

Several of the authors whom I cited earlier and indeed some others offer a positive answer to this question. For instance, in the previously cited passage Glannon expresses the belief that data obtained from *functional* neuroimaging could excuse a person of responsibility by showing that they lacked a responsibility-relevant capacity. Raine and Yang's previously cited comments about the responsibility of delinquent, criminal, violent and psychopathic populations were also informed in part by data obtained from *functional* neuroimaging studies. Similarly, James Blair's ruminations about psychopaths' responsibility are also derived from *functional* neuroimaging data; for instance, data obtained from *functional* neuroimaging studies is the basis of his claim that the capacity for intention is undermined in the psychopath because of an "increased risk for reactive aggression" ([39]:153), and that is obviously also the case in his claim that "[t]he amygdala and vmPFC *dysfunction* ... will seriously impair the decision making in psychopathy" ([39]:154, emphasis added).<sup>17</sup> Finally, the idea that *functional* neuroimaging might be capable of revealing people's reduced-responsibility is also evident in Claudia Pinto's article about the man who "was a schoolteacher, a husband, a father[, and t]hen he became a pedophile preoccupied with sex" in which she writes that "murderers, as a group, had lower glucose metabolism in the

<sup>17</sup> Blair's claim that "[t]wo important capacities are necessary for successful [moral] socialization. First, the individual must be capable of finding the distress of others aversive....Second, the individual must be capable of performing stimulus-reinforcement learning", commits him to a capacitarian account of the conditions of responsibility ([38]:327).

prefrontal cortex, showing diminished activity in brain areas which normally function to inhibit aggressive impulses” [43].

A range of people think that *functional* neuroimaging *could* in principle be used to help us assess a person’s responsibility by revealing their mental capacities.

### The Worry

However, it could be argued that a modal fallacy similar to the one that was described earlier might also be committed if *functional* neuroimaging is used to assess people’s capacities.<sup>18</sup> To see how this might happen, consider the following example. Suppose, that we are trying to establish whether Jane and John are capable of (i.e. whether they have the *capacity* for) doing maths, and to simplify matters suppose also that mathematical ability in people who have it is implemented in a single part of the brain called Brodmann Area Maths (BAM).<sup>19</sup> We put Jane and John into fMRI scanners, present each with a number of equations, and ask them to determine whether those equations are correct or not. Now, as it happens, Jane was never taught maths, and so although she is enthusiastic she ultimately performs badly—Jane never developed the brain mechanisms which one

<sup>18</sup> To recap, one commits a modal fallacy when one moves from an observation of the form “X doesn’t happen” (i.e. an observation about *behaviour*) to a conclusion of the form “X can’t happen” (a claim about *capacity*), and I argued above that this is precisely the sort of move which is made within the behavioural approach because claims about people’s incapacities are derived from observations about what sorts of behaviour they do not exhibit. However, I then argued that by using the neuroimaging approach we could avoid committing a modal fallacy since here claims about people’s capacities are derived from observations of their brain mechanisms.

<sup>19</sup> In actual fact, no *single* area of the brain is associated with mathematical reasoning, but rather a *number of different areas* have been implicated including the “angular gyrus, left parietal and prefrontal cortex” ([44]:168). This is due to the fact that there seems to be at least two different kinds of mathematical reasoning: “*Exact arithmetic* puts emphasis on language-specific representations and relies on a left inferior frontal circuit also used for generating associations between words.... *Approximate arithmetic*, in contrast, shows no dependence on language and relies primarily on a quantity representation implemented in visuo-spatial networks of the left and right parietal lobes” ([45]:973, emphasis added). Never the less, to simplify this example, I talk as if there were just one area of the brain—namely, the fictional Brodmann Area Maths (BAM)—that is the seat of mathematical ability.

needs to have in order to understand let alone to do maths (her BAM does not have a capacity to do maths), and so the fMRI scan shows below-average activation in her BAM. On the other hand, John teaches graduate courses in maths at Big State University, but today he is feeling unmotivated, and so he only pretends to assess the equations while actually thinking about something else—John does not use his BAM, and so not surprisingly the fMRI images of John’s brain look similar to those of Jane’s brain. However, although the fMRI scans reveal below-average levels of activation in *both* Jane and in John’s BAM, this surely should not be taken as evidence that both of them lack the capacity to do maths, since *ex hypothesi* we know that John does have the capacity to do maths but it’s just that he wasn’t exercising that capacity when we scanned his brain!<sup>20</sup>

Coming back to the case at hand, even if we noted that the brains of people who consistently perform badly in various moral judgment tasks also show consistently reduced patterns of activation in some area of the brain that would normally be used by subjects who perform well in those tasks—for instance, in “the amygdala and rostral anterior cingulate cortex/vmPFC” ([39]:151)—why should we suppose that this reveals an *incapacity* (i.e. that those parts of their brain *can not* function at higher levels of activation) rather than merely showing that those people *do not use* a capacity which, for all we know, they might actually possess? How could we distinguish whether such people are more like Jane who non-culpably<sup>21</sup> lacks the capacity for maths (i.e.

<sup>20</sup> In this example I am working under the simplifying assumption that with greater exertion of mathematical skill come higher levels of BAM activation, and that with lesser exertions come lower levels of BAM activation. But as a person studies more maths, the connections between the neurons in their BAM might become optimised through dendritic pruning, and this may result in increasingly *lower* levels of activation in their BAM the more proficient they become at doing maths. However, I could have made the same point while assuming that there is an inverse (or some other more complex) relationship between skilled exercise of mathematical ability and BAM activation levels, and so I ask the reader to bear with this assumption. I am grateful to Jonathan Opie from the University of Adelaide for pointing this out to me.

<sup>21</sup> The point of the “non-culpably” clause is to cut short the potential objection that perhaps these people are responsible for their incapacity.

perhaps they too non-culpably lack the capacity to correctly assess moral judgment tasks), or whether they are more like John (perhaps they are lazy,... perhaps they just do not put in the same effort as the rest of us do to figure out what is the right thing to do in various situations)?<sup>22</sup> The point is that if *functional* neuroimaging was our exclusive tool of choice then we could *not* distinguish someone who possessed a capacity but failed to use it from someone who simply lacked that capacity altogether. As Walter Glannon points out, just like “[b]ehavioral criteria ... do not ... help us to distinguish between having some ... capacity and failing to exercise it ... and lacking the capacity [altogether, s]imilar remarks apply to brain imaging studies” ([46]:161).

The source of this problem is that by definition, a *functional* neuroimaging scan is only designed to reveal parts of the brain which are used in the performance of a particular cognitive task, and so we should not expect to find any differences in the scans of those people who *lack a given capacity and that is why they do not use it* and those who *possess that capacity but never the less do not use it*, because in both cases the area of the brain associated with that capacity will not get used. Although *structural* scans of their respective brains may indeed look different—they might for instance reveal that one person possesses the requisite brain mechanisms whereas the other person does not—we should not expect *functional* scans to look any different. Unlike *structural* neuroimaging which reveals the mechanisms or “hardware” in which a person’s capacities are implemented, *functional* neuroimaging reveals only which parts of the brain a person was using when they allegedly attempted to perform a particular cognitive task. However, this means that when conclusions about a person’s *capacities* are derived from a *functional* scan of their brain, then those conclusions *will* still be derived from observations about *behaviour*—brain behaviour—and so it would seem that when *functional*

<sup>22</sup> Exactly the same question could be asked with respect to claims about *volitional* capacities that are derived from *functional* neuroimaging data, except here we would ask something like “How could we distinguish whether such people are more like Jane (i.e. perhaps they too lack the capacity to control their angry impulses), or whether they are in fact more like John (perhaps they are just lazy and they just do not put in the same effort as the rest of us do to restrain ourselves)?”

neuroimaging data is used to support claims about capacities then a modal fallacy *will* after all be committed!

Finally, it is important to note that this worry does not just attach to *current* functional neuroimaging techniques, but rather it will attach even to *future* functional neuroimaging technology; if the future technology reveals only *functional* data then according to the above arguments one will have to commit a modal fallacy in order to derive claims about capacities from that data. Thus, what is it that justifies the transition from the observation (obtained through a functional neuroimaging scan) that some part of a person’s brain consistently shows lower-than-average levels of activation (i.e. reduced function) to the claim that this part of their brain can not (i.e. that it lacks the capacity to) operate at higher levels of activation?

### The Solutions

The transition that is made when functional neuroimaging data is cited to support claims about people’s capacities is an inference from an observation about behaviour (in this case brain behaviour) to a conclusion about capacities (how that brain can not behave), and the worry described above is that this inference is *not justified*. However, I think that there are at least two ways to justify this transition and thus to show that this inference *is* warranted after all—one to do with automatic responses, and the other to do with theories to the best explanation—which deflect the worry that a modal fallacy is committed when this inference is made, and I will now say something about each of these justifications.

#### *Automatic Responses*<sup>23</sup>

The first justification looks something like this.

If the areas of the brain in which some part of the capacity for sound moral judgment is normally implemented—e.g. the amygdale—are ones which are meant to activate *automatically* whenever subjects are exposed to certain kinds of stimuli (e.g. when they

<sup>23</sup> I am grateful to Philip Gerrans from the University of Adelaide for pointing out this justification for deriving conclusions about capacity from observations about consistently reduced function.

are shown images of angry and sad faces or when they hear emotionally charged words), then we could legitimately surmise that a person indeed has reduced capacity for sound moral judgment if those parts of their brain consistently show reduced levels of activation when they are presented with the right kind of stimuli. A similar move is made in the diagnosis of “cortical visual impairment”, where the consistent lack of activity in a person’s occipital lobe when they are presented with appropriate visual stimuli can be taken as evidence that this part of their brain lacks the capacity to process visual information.<sup>24</sup> Hence, similarly, if the capacity for sound moral judgment also requires some *automatic* responses in certain areas of the brain when subjects are exposed to the right kinds of stimuli, then when we consistently observe a significantly reduced level of activation in those areas of the brain upon exposure to those stimuli, we should also be entitled to conclude that this person has reduced capacity for sound moral judgment.<sup>25</sup>

And indeed, it has been argued that there *are* certain automatic responses that can be observed when normal subjects are exposed to moral cognition inducing stimuli and which are required for sound moral judgment. For instance, when normal subjects are presented with pictures of fearful or sad faces, they automatically show an elevated skin conductance response [49] and increased activity in their amygdala [39], and it is hypothesized that both of these automatic responses (and a horde of others) are required for sound moral judgment [26]; without these automatic responses subjects either fail to perceive important moral cues, or they simply fail to engage in the right kind of cognitive processing, and hence such people’s moral judgments can indeed fail to be sound.<sup>26</sup>

Thus, the first sort of case in which we will be justified in drawing a conclusion about reduced

capacity from an observation about consistently reduced levels of neural activation, is when the missing activation was meant to be an automatic response to the given stimuli. Put another way, if exposure to the given stimuli should always raise the level of activation in a specific area of every subject’s brain, but yet that area of *this* subject’s brain did not show that response despite the fact that they were presented with the right stimuli, then we may justifiably conclude (i.e. without committing a modal fallacy) that a mechanism which normally generates an automatic response is broken and hence that the subject does indeed have a reduced capacity to respond in the given manner.

### *Theory to the Best Explanation*<sup>27</sup>

The derivation of claims about reduced capacity from *functional* neuroimaging data can also be justified in another way.

If my car consistently fails to start on cold mornings but yet on warm mornings it starts without any noticeable problems, then I will start to suspect that there is something mechanically wrong with it—for instance, that the spark plugs are worn out and that on cold mornings when metal warps the gap between the spark plugs’ electrodes becomes too large for the electricity to successfully arc across them and to ignite the petrol in the engine’s cylinders. Similarly, if buses consistently run late on Monday mornings then I will start to suspect that there is something about how people behave on the first day of the working week that causes this phenomenon—for instance, that congestion is caused when everyone turns up at the same time on the roads, at bus and tram stops and at train stations, and that this congestion slows everything down. In both of these cases, my observation of some consistent anomaly leads me to posit the existence of some as-yet-unobserved mechanism that explains the observed phenomenon, and once I have this mechanism I will then be able to validly deduce some further claims about capacities—for instance, about my car’s reduced capacity to start on cold mornings, and about

<sup>24</sup> For discussion of the use of *functional* neuroimaging techniques in the diagnosis of cortical visual impairment, see Isaac Silverman et al. [47] who used *hexamethylpropylene amine oxime* in their SPECT studies to obtain functional data, or William Good et al. [48] who suggest that fMRI might also be used to obtain such functional data.

<sup>25</sup> Though please note my comments in note 19 above.

<sup>26</sup> Also see Glannon’s useful summary of relevant empirical findings ([46]:159–60).

<sup>27</sup> I am grateful to Gert-Jan Lokhorst of TU Delft for suggesting this second justification for deriving conclusions about capacity from observations about consistently reduced activation.

the public transport system's reduced capacity to run on time on Monday mornings.

I think that something analogous also happens in the case at hand, where the claim that someone lacks the capacity for sound moral judgment (or some necessary constituent of that capacity) is also intended as an inference to the best explanation—i.e. the observation of a consistently reduced level of activity in a certain part of that person's brain when that person is exposed to stimuli which should have led to an increased level of activation leads us to hypothesize that there must be some as-yet-unobserved structural anomaly in their brain which explains the consistently observed lower level of activity, and since no modal fallacy is committed when claims about reduced capacity are derived from claims about structure—it doesn't take much to imagine how a lack of the appropriate brain mechanism may result in a lack of the related capacities—similarly, no modal fallacy will be committed when claims about capacities are derived from such hypotheses about structural abnormalities. Put another way, a transition seems to be made in the case at hand from an observation about the *functional* profile of the person's brain (derived from a functional neuroimaging scan) to a hypothesis about the *structural* features of their brain, and the claim about reduced capacity is then derived from the claim about this hypothesized structural anomaly and not directly from the functional data. Naturally, another objection may now be raised—namely, that the claim about the person's reduced capacity is now derived from something that is only postulated to exist but which has not yet been shown to exist—however this is not the same as the modal fallacy problem.

Hence, in the second instance, the transition from the observation of a consistently reduced level of activation in some part of a person's brain to the claim that this person has a reduced capacity for sound moral judgment, can also be justified as an attempt to offer the best available explanation for the observed phenomenon. The “best explanation for the low brain activity as measured functionally is that there is a structural anomaly that accounts for the [observed] functional problem”,<sup>28</sup> and no modal fallacy is committed when conclusions about incapacity are

deduced from claims about structural features of the brain.

### Section Summary

Thus, there are two ways to derive claims about reduced capacity from functional neuroimaging data without committing a modal fallacy. Firstly, no modal fallacy will be committed if the activity that is missing was meant to be an automatic response to some specific stimuli. Secondly, if the claim about reduced capacity is derived not directly from the functional neuroimaging data itself, but rather indirectly from a hypothesis about some as yet unobserved structural anomalies—a hypothesis which explains that functional data—then again no modal fallacy will be committed since claims about reduced capacities can be derived from structural data without committing a modal fallacy.

As an aside, similar replies to the modal fallacy problem might also be available to proponents of the behavioural approach to the individualized assessment of responsibility.<sup>29</sup> For instance, if we have good reason to believe that all people who possess the capacity for sound moral judgment will always wince whenever they see a photo of someone in a distressing situation and that this is an involuntary response which can not be consciously suppressed by the subject who is being tested, then when we observe someone who fails to exhibit this automatic response we will also be justified in drawing the conclusion that this subject has reduced capacity for making sound moral judgments. Similarly, if proponents of the behavioural approach can posit a plausible psychological mechanism which explains the observed lack of some behaviour, and then derive their claims about a specific person's reduced cognitive or volitional capacity from claims about this as-yet-unobserved psychological mechanism, then that too may enable them to avoid committing a modal fallacy. However, although this suggests that the modal fallacy problem may not be as damaging to the behavioural approach as I initially suggested (with support from McSherry and Sapolsky), it does still mean that for the behavioural approach to provide useful evidence we must use more complex tests—i.e. ones which

<sup>28</sup> I thank Eddy Nahmias of Georgia State University for this helpful way of putting my point.

<sup>29</sup> I thank the editor of Neuroethics for setting me straight on this point.



do not permit the subject whose responsibility is being tested to intentionally interfere in the results—and that we should not rely too heavily on biographical data obtained from other’s stories about the person whose responsibility is being assessed since my original comments in this regard still stand.

## Conclusion

In *Roper v. Simmons* [1] the Supreme Court of the United States spared Simmons’ life when it ruled that Simmons’ execution would be unconstitutional because “[t]he Eighth and Fourteenth Amendments forbid imposition of the death penalty on offenders who were under the age of 18 when their crimes were committed”. Presumably, age matters to responsibility because of its relationship to maturity, and maturity matters to responsibility because usually with greater maturity come greater mental capacities—of most interest to us here, greater cognitive and volitional capacities. However, people mature at different rates, and so we oughtn’t rely on age to be an accurate indicator of a person’s true capacities and of how responsible they truly are. Given what hangs on responsibility assessments—i.e. the type and degree of punishment to which an accused person will be subjected—such assessments should be as precise and accurate as possible, and the way to get greater precision and accuracy is to find a more individualized way of assessing people’s responsibility, rather than using a person’s age as a proxy for the degree of their responsibility.

One way to individually assess people’s responsibility is through what I referred to as the behavioural approach. However, this approach may still be dreadfully inaccurate if it relies on the veracity of those who are being tested and on the veracity of those who are asked to characterize that person, and furthermore unless such behavioural approaches are properly designed then they will run afoul of the modal fallacy problem that was described above. For these reasons we should at least supplement these behavioural approaches to the individualized assessment of responsibility with something else.

I have argued that neuroimaging offers us such a supplement—a promising way to individually assess people’s responsibility. Although neuroimaging is no “silver bullet”—i.e. at best it only answers *some* of

the questions that need to be answered in order to assess a person’s responsibility—and although it also comes with its own suite of problems, never the less in time some of those problems will most probably be overcome, and the remainder of these problems is also encountered by the behavioural approach. Thus, given the foregoing arguments, it is now my contention that in the future—i.e. once science and technology have progressed sufficiently—neuroimaging may indeed be helpful in this regard. Neuroimaging can free us of our current reliance upon such rough-and-ready rules of thumb as the claim that a person’s responsibility is in part determined by their age—rules of thumb which carried the day in *Roper v. Simmons* [1]—and given what is at stake for those whose responsibility is currently assessed in this manner and the fact that this rule of thumb is ultimately inaccurate, this can only be a good thing.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Roper v. Simmons. 2005. Donald P. Roper, Superintendent, Potosi Correctional Center, Petitioner v. Christopher Simmons, <http://laws.findlaw.com/us/543/551.html>, accessed 2009109.
2. Herndon, J., and S.P. Waxman. 2004. *Brief for respondent in Roper v Simmons [2005]*. St. Louis, MO: i-xiv, 1-50, 1a-21a.
3. ABA-1. “Christopher Simmons-Juvenile Death Penalty”, <http://www.abanet.org/crimjust/juvjus/simmons.html>, accessed 20081021.
4. Wikipedia. “Roper v. Simmons” [http://en.wikipedia.org/wiki/Roper\\_v.\\_Simmons](http://en.wikipedia.org/wiki/Roper_v._Simmons), accessed 20081021.
5. Perry, S.R. 2000. Loss, agency, and responsibility for outcomes: Three conceptions of corrective justice. In *Philosophy of law*, eds. J. Feinberg, and J. Coleman, 6546–559. Belmont: Wadsworth/Thompson Learning.
6. Hart, H.L.A. 1968. *IX. Postscript: Responsibility and retribution. Punishment and responsibility*. 210–237. Oxford: Clarendon.
7. Kutz, C. 2004. Chapter 14: Responsibility. In *Jurisprudence and philosophy of law*, eds. J. Coleman, and S. Shapiro, 548–587. Oxford: Oxford University Press.
8. Vincent, N. 2006. *Responsibility, Compensation and Accident Law Reform. Discipline of Philosophy, Faculty of Humanities and Social Sciences*. Adelaide: University of Adelaide. Doctor of Philosophy <http://digital.library.adelaide.edu.au/dspace/handle/2440/39507>.
9. Fischer, J.M., and M. Ravizza. 1998. Conclusion. In *Responsibility and control: A theory of moral responsibility*,

- eds. J.M. Fischer, and M. Ravizza, 240–259. Cambridge: CUP.
10. Dennett, D.C. 1998. On giving libertarians what they say they want. In *Philosophy then and now*, eds. N.S. Arnold, T.M. Benditt, and G. Graham, 119–130. Malden: Blackwell.
  11. Wolf, S. 1987. Sanity and the metaphysics of responsibility. In *Responsibility, character and the emotions: New Essays in moral psychology*, ed. F. Shoeman, 46–62. New York: Cambridge University Press.
  12. Wolf, S. 2001. The reason view. In *Agency and responsibility: Essays on the metaphysics of freedom*, ed. L.W. Ekstrom, 205–226. Boulder: Westview.
  13. Sapolsky, R.M. 2004. The frontal cortex and the criminal justice system. *Philosophical Transactions of the Royal Society of London* 359: 1787–1796.
  14. Roskies, A. 2006. Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences* 10(9): 419–423.
  15. Morse, S.J. 2006a. Brain overclaim syndrome and criminal responsibility: A diagnostic note. *Ohio State Journal of Criminal Law* 3: 397–412.
  16. Churchland, P.S. 2006. The big questions: Do we have free will. *New Scientist* 2578: 42–45. (18-Nov).
  17. Glannon, W. 2005. Neurobiology, neuroimaging, and free will. *Midwest Studies in Philosophy* XXIX: 68–82.
  18. Morse, S.J. 2006b. Moral and legal responsibility and the new neuroscience. In *Neuroethics: Defining the issues in theory, practice, and policy*, ed. J. Illes, 33–50. Oxford: Oxford University Press.
  19. Vincent, N. 2008. Responsibility, dysfunction and capacity. *Neuroethics* 1(3): 199–204.
  20. Levy, N. 2007. The Responsibility of the psychopath revisited. *Philosophy, Psychiatry & Psychology* 14(2): 129–138.
  21. Fischer, J.M. 2005. Reply: The free will revolution. *Philosophical Explorations* 8(2): 145–156.
  22. McKenna, M. 2000. Assessing reasons-responsive compatibilism. *International Journal of Philosophical Studies* 8 (1): 89–124.
  23. McKenna, M. 2001. Book review: Fischer and Ravizza's "Responsibility and control". *The Journal of Philosophy* 98 (2): 93–100.
  24. McKenna, M. 2005. Reasons reactivity and incompatibilist intuitions. *Philosophical Explorations* 8(2): 131–143.
  25. McSherry, B. 2004. Criminal responsibility, 'fleeting' states of mental impairment, and the power of self-control. *International Journal of Law and Psychiatry* 27: 224–257.
  26. Greene, J., and J. Haidt. 2002. How (and where) does moral judgment work. *Trends in Cognitive Sciences* 6(12): 517–523.
  27. Gerrans, P., and J. Kennett. 2006. Introduction: Is cognitive penetrability the mark of the moral. *Philosophical Explorations* 9(1): 3–12.
  28. Greene, J., L.E. Nystrom, et al. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.
  29. Matthews, S. 2004. Failed agency and the insanity defence. *International Journal of Law and Psychiatry* 27: 413–424.
  30. Raine, A., and Y. Yang. 2006. Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience* 1(3): 203–213.
  31. Goodenough, O.R., and K. Prehn. 2004. A neuroscientific approach to normative judgment in law and justice. In *Law & the brain*, eds. S. Zeki, and O.R. Goodenough New York: Oxford University Press.
  32. Eastman, N., and C. Campbell. 2006. Neuroscience and legal determination of criminal responsibility. *Nature Reviews Neuroscience* 7: 311–318. (April).
  33. Mobbs, D., H.C. Lau, et al. 2007. Law, responsibility, and the brain. *PLoS Biology* 5(4): 693–700.
  34. Logothetis, N.K. 2008. What we can do and what we cannot do with fMRI. *Nature* 453: 869–878. (12 June 2008).
  35. Tallis, R. 2007. Why blame me? It was all my brain's fault—The dubious rise of 'neurolaw'. *The Times*, 24 October 2007.
  36. Morse, S.J. 2004. (Thursday, 9th September 2004). Session 1: Neuroscience, Brain, and Behavior VI: Neuroscience and the Law, from <http://www.bioethics.gov/transcripts/sep04/session1.html>, accessed 200711128.
  37. Reimer, M. 2008. Psychopathy without (the Language of) Disorder. *Neuroethics* 1(3): 185–198.
  38. Blair, R.J.R. 2007. Aggression, psychopathy and free will from a cognitive neuroscience perspective. *Behavioral Sciences and the Law* 25: 321–331.
  39. Blair, R.J.R. 2008. The cognitive neuroscience of psychopathy and implications for judgments of responsibility. *Neuroethics* 1(2): 149–157.
  40. Maibom, H.L. 2008. The mad, the bad, and the psychopath. *Neuroethics* 1(3): 167–184.
  41. Jones, O.D. 2004. Law, evolution and the brain: Applications and open questions. In *Law & the brain*, eds. S. Zeki, and O.R. Goodenough, 57–75. New York: Oxford University Press.
  42. van Inwagen, P. 2001. The incompatibility of free will and determinism. In *Agency and responsibility: Essays on the metaphysics of freedom*, ed. L.W. Ekstrom, 17–29. Boulder: Westview.
  43. Pinto, C. 2003. Putting the brain on trial. Media General News Service, <http://www.rifters.com/real/articles/brainontrial.htm>, accessed 20080706.
  44. Krendl, A.C., J.A. Richeson, et al. 2008. The negative consequences of threat: A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science* 19(2): 168–175.
  45. Dehaene, S., E. Spelke, et al. 1999. Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science* 284: 970–974.
  46. Glannon, W. 2008. Moral responsibility and the psychopath. *Neuroethics* 1(3): 158–166.
  47. Silverman, I.E., S.L. Galetta, et al. 1993. SPECT in patients with cortical visual loss. *The Journal of Nuclear Medicine* 34(9): 1447–1451.
  48. Good, W.V., J.E. Jan, et al. 2001. Recent advances in cortical visual impairment. *Developmental Medicine & Child Neurology* 43(1): 56–60.
  49. Damasio, A.R. 1995. *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Harper Perennial.