

Restoring Responsibility: Promoting Justice, Therapy and Reform Through Direct Brain Interventions

Nicole A Vincent

Published online: 25 May 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Direct brain intervention based mental capacity restoration techniques—for instance, psycho-active drugs—are sometimes used in criminal cases to promote the aims of justice. For instance, they might be used to restore a person’s competence to stand trial in order to assess the degree of their responsibility for what they did, or to restore their competence for punishment so that we can hold them responsible for it. Some also suggest that such interventions might be used for therapy or reform in criminal legal contexts—i.e. to make non-responsible and irresponsible people more responsible. However, I argue that such interventions may at least sometimes fail to promote these responsibility-related legal aims. This is because responsibility hinges on other factors than just what mental capacities a person has—in particular, it also hinges on such things as authenticity, personal identity, and mental capacity ownership—and some ways of restoring mental capacity may adversely affect these other factors. Put one way, my claim is that what might suffice for the restoration of competence need not necessarily suffice for the restoration of responsibility, or, put another way, that although responsibility indeed tracks mental capacity it may not always track restored mental capacities.

Keywords Capacitarianism · Responsibility · Mental capacity restoration · Direct brain intervention · Competence for trial · Competence for punishment · Competence for release · Justice · Therapy · Reform · Characterianism

An early draft of this paper was presented on 12th April 2011 at the *TILTING Perspectives 2011: Technologies on the stand* conference at Tilburg University in The Netherlands, and it was published in the conference proceedings under the title “Capacitarianism, Responsibility and Restored Mental Capacities”.

N. A Vincent
Philosophy Department, Macquarie University, Sydney, NSW, Australia

N. A Vincent (✉)
Philosophy Section, TU Delft, Delft, The Netherlands
e-mail: nicole.vincent@mq.edu.au

Introduction

The capacitarian idea that responsibility tracks mental capacity underlies much of our thinking about responsibility. For instance, mental capacity assessments inform whether someone is seen as a fully responsible person, what responsibilities they can be expected to observe, their degree of responsibility for what they did, and whether they can be expected to take responsibility for what they did and be held responsible for it. But what happens when mental capacities are restored using medical treatments? Could such treatments help us to assess the responsibility of someone who becomes mentally ill after committing their crime, to expect them to take responsibility for what they did, to hold them responsible for it, to make them fully responsible and maybe even less irresponsible? Are we entitled to assume that responsibility will keep on tracking such restored mental capacities, or that responsibility can be restored this way?

After (1) demonstrating the prominent role that capacitarianism plays in lay, legal and philosophical thinking, and (2) showing how some legal uses of mental capacity restoration techniques presuppose an extension of capacitarianism, I then (3) identify objections to this extension of capacitarian thinking—namely, that some ways of restoring mental capacity might adversely affect other factors which also bear on responsibility, namely authenticity, personal identity and mental capacity ownership. Following (4) a brief commentary on why these objections tend to arise in particular when direct brain intervention based mental capacity restoration techniques are used to promote the aims of justice, therapy and reform, I then conclude that (5) what might suffice for the restoration of competence need not necessarily suffice for the restoration of responsibility, or, put another way, that although responsibility indeed tracks mental capacity it may not always track restored mental capacity.

Capacitarianism Underlies Much of Our Thinking About Responsibility

The “capacitarian” idea that responsibility co-varies with or “tracks” mental capacity underlies much lay, legal and philosophical thinking about responsibility.

In *lay* contexts responsibility is often thought to require such things as the ability to perceive the world without delusion, to think clearly and rationally, to guide our actions by the light of our judgments, and to resist acting on mere impulse. This is, for instance, why children, the senile, and the mentally ill are thought to be less than fully responsible for what they do (i.e. because they lack the right kind and/or degree of mental capacity), why children can acquire more and/or greater responsibilities as they grow up (i.e. because their mental capacities develop as they mature), and how responsibility is reinstated on recovery from mental illness (i.e. because the needed mental capacities are recovered).

Identical sentiments are expressed in the *legal* context by the idea that responsibility requires certain “cognitive” and “volitional” mental capacities. H. L. A. Hart suggests that “[t]he capacities in question are those of understanding, reasoning, and control of conduct: the ability to understand what conduct legal rules or morality require, to deliberate and reach decisions concerning these requirements, and to conform to decisions when made” (1968:227). The American Law Institute’s Model Penal Code states that “a person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to *appreciate* the criminality of his conduct [i.e. cognitive capacity] or to *conform* his conduct to the requirements of the law [i.e. volitional capacity]” (MPC 1962:§4.01, my emphasis; also see Robinson and Dubber

2007:339). The capacitarian approach provides an effective framework for understanding how legal defences such as insanity and automatism diminish responsibility—namely, by highlighting deficits in the mental capacities required for full responsibility, and challenging either the claim that the person’s behaviour was an action (and thus that it can be attributed to them *qua* action) or the claim that they can be blamed for having violated their responsibilities in acting as they did (because they non-culpably lacked the capacity to act as they ought to have acted). Finally, many recent attempts to introduce neuroscience and behavioural genetics into courtrooms are also best understood in precisely this way—i.e. they are introduced to help make assessments of mental capacity which in turn inform assessments of responsibility (Vincent 2011a:80–84, 2010).

Mental capacity assessments also play a central role in the legal practice of expecting defendants to take responsibility for what they did (in the sense of accounting for their actions at trial, and acknowledging their guilt), and of holding defendants responsible (in the sense of punishing them for it). For instance, in *Connecticut v. Kenneth Curtis* (1999) a defendant who shot himself in the head after first shooting and killing his estranged girlfriend was initially found to be incompetent to stand trial, and thus he was initially neither tried nor punished. However, years later he made such an impressive recovery that he even enrolled in college, and upon subsequent re-assessment he was found competent to stand trial where he pled guilty to manslaughter and received a 20 year prison sentence. In *Ford v. Wainwright* (1986) the court held that it was impermissible to punish defendants who are insane (even if they were not insane at the time when they committed the crime) because this would serve no penological goals. And similar arguments were also cited in *Atkins v. Virginia* (2002) where the court held that it was impermissible to punish mentally retarded defendants. The idea that responsibility tracks mental capacity clearly plays a key role in the legal practices of holding people responsible and of expecting them to take responsibility for what they did.

Furthermore, consider the role that mental capacity assessments play in parole boards’ deliberations. One question a parole board must consider is whether the person under assessment has reformed sufficiently such that it is now safe to release them back into society—i.e. whether we have sufficient evidence that they are no longer irresponsible and that they can be trusted to stay out of trouble. Another question concerns whether the person in question has sufficient mental capacities to be a fully responsible person in the sense that on release from prison they could fend for themselves, and, should the need arise, be held responsible (in the relevant senses of the expression) for any subsequent wrongdoing.¹

Finally, the idea that responsibility tracks mental capacity is also central to much *philosophical* thinking. For instance, Aristotle wrote that “feelings and actions ... that are involuntary receive pardon” and that “[a]ctions are regarded as involuntary when they are performed under *compulsion* or through *ignorance*” (Aristotle 1976:111, emphasis added)—i.e. with compromised volitional or cognitive capacity. And since Aristotle, this view has also been endorsed and developed by numerous other philosophers (e.g. Dennett 1984; Wallace 1994; Glannon 2002; Sher 2009). Notably, capacitarianism is at the centre of John Fischer’s and Mark Ravizza’s compatibilist defence of responsibility from the alleged threat of determinism. Fischer and Ravizza contend that the reason why we retain

¹ What I list here as two separate questions— first, whether the person concerned is no longer irresponsible, and second, whether they are now fully responsible—might seem like two ways of asking the same question. However, as I explain in section “[Worries About Using Medical Treatments for Therapy and Reform](#)” below, there are good reasons to distinguish these questions from one another.

the intuition that agents in Frankfurt-style counter examples (Frankfurt 1969:835–836) are still responsible for what they did even though they could not have done otherwise is because they possessed “guidance control” (Fischer and Ravizza 1998:34–41). On their account, a person has guidance control when their actions issue from their own² “moderately reasons-responsive mechanisms”, or when via “tracing” we establish that they are responsible for the fact that their mechanism was not moderately reasons-responsive (1998:49–51). Mechanisms are moderately reasons-responsive when they are “regularly receptive” and “weakly reactive” to reasons. And a mechanism is weakly reactive to reasons if it reacts to those reasons in at least a small set of possible worlds, even though it may not react to those reasons in this actual world (1998:69–85). Simplifying somewhat, on Fischer and Ravizza’s account a person has guidance control and is thus responsible for what they do when their actions issue from mechanisms in virtue of which they possess an appropriate measure of the cognitive and volitional mental capacities which are required for moral agency. Reasons-responsiveness is a feature of certain brain mechanisms in virtue of which one can comprehend what one is doing and what one ought to do/not do, and be moved appropriately to action on the basis of that comprehension; and the moderateness component captures the idea of these brain mechanisms having *varying degrees of capacity* to do those things.

As I argue elsewhere (Vincent 2010, 2011b), and as the above discussion hopefully makes evident, the word “responsibility” refers to a range of different though related concepts and practices. And what I hope to have demonstrated above (albeit briefly) is that in lay, legal and philosophical contexts we often presuppose the existence of a certain kind of relationship between claims about responsibility in its various senses and claims about the mental capacity of those to whom the responsibility claims pertain—namely, that responsibility co-varies with or tracks mental capacity. Mental capacity evaluations inform our views on whether someone is a *fully responsible person*, their degree of *responsibility for something that they did*, and whether they can be *held responsible* for what they have done in the sense of standing trial and being answerable.³ Furthermore, mental capacity evaluations also inform what sorts of *responsibilities it is reasonable to impose* onto people, and whether we can *hold them responsible* by imposing punishment and liability onto them.

Capacitarianism and Restored Mental Capacities

The previous section highlighted the pervasive role that capacitarianism plays in our thinking about responsibility, and I take this pervasiveness to indicate the strength of our commitment to the idea that *responsibility tracks mental capacity*. In this section I will argue that actual and proposed uses of mental capacity restoration techniques in legal contexts often presuppose an extension of this capacitarian thinking—namely, that

² This refers to Fischer and Ravizza’s ownership condition (1998:207–239). Given that this paper’s focus will ultimately shift to this condition, I defer discussing its significance till later.

³ Please note the contrast between *this* use of “holding responsible”, which is about accountability (e.g. van de Poel 2011:39 and 47), and the sense of “holding responsible” mentioned in the *next sentence*, which relates to how a convicted person should be treated on account of their blameworthiness—e.g. punished or made liable (e.g. see Vincent 2011b:16 and 18).

responsibility also tracks *restored* mental capacities, or, put another way, that responsibility can be *restored* through mental capacity restoration.

Mental Capacity Restoration and Justice

Consider the way in which mental capacity restoration techniques are already being used to bring criminals to *justice*. Criminals who develop mental disorders after committing crimes may lack competence to stand trial or to be punished, and this prevents them from taking responsibility and us from trying them to assess their responsibility and subsequently from holding them responsible for what they did. However, such mental disorders can nowadays often be treated with medications or other medical procedures (Lekovic 2008).

For instance, defendants are sometimes involuntarily treated with anti-psychotic drugs to restore their competence to stand trial. In *Riggins v. Nevada* (1992) the court held that it was generally permissible to administer anti-psychotic drugs that make defendants competent to stand trial, as long as this is medically appropriate and the least intrusive means available. In *Sell v. United States* (2003) the US Supreme Court clarified this position by ruling that it is permissible to forcibly medicate a defendant for the *sole* purpose of making them competent to stand trial, as long as the treatment would most likely be effective, no better medical alternatives are available, and important state interests are at stake (e.g. bringing criminals to justice). And defendants are sometimes treated with anti-psychotic drugs to restore their competence for punishment—e.g. in *Singleton v. Norris* (2003) anti-psychotic drugs were forcibly administered to a condemned inmate to make him competent for execution by lethal injection. But even if we do not endorse capital punishment—for instance, I do not—we might still endorse the more moderate idea that inmates who develop mental disorders while serving a prison sentence should receive medical treatment, and not only because prisoners retain a right to adequate medical care, but to ensure that the retributive aim of incarceration (one of the aforementioned penological goals cited in *Ford v. Wainwright* (1986) above) is achieved.

These examples express the underlying idea that by restoring people's competence to stand trial and to be punished we can restore their ability to *take responsibility*, and our ability to assess their degree of *responsibility for what they did* and to *hold them responsible* for it—i.e. that responsibility also tracks *restored* mental capacities.

Mental Capacity Restoration for Therapy and Reform

Now consider the use of mental capacity restoration techniques in the pursuit of broadly *therapeutic* and *reformatory* legal aims.

Some authors write that at least a portion of criminal behaviour may be caused by mental disease or disorder (Sapolsky 2004; Tancredi 2005), and so they suggest that at least these people should receive medical treatments that restore their mental capacities rather than being punished. It is for precisely this kind of reason that repeat sex offenders are sometimes given drugs like cyproterone acetate, a powerful antiandrogen that allegedly helps them to regain self-control by reducing their sex drive (Bradford and Pawlak 1993). As Hank Greely puts it, such drugs “change the convict's behavior, reducing his previously compulsive interest in sex from a scream to a whisper” (2009:177). This is also the thinking behind trials currently under way in Australia (Macey 2010) in which convicted violent criminals are treated with selective serotonin reuptake inhibitors (SSRIs) which

Khalifa et al. (2010) report can help people control their anger⁴—the aim is, yet again, to restore their capacity for self-control.⁵ With similar aims, though using different techniques, neurosurgeons have reported increasing success rates and diminishing rates of complications and side effects in treating severe aggression through the surgical removal or lesioning of the amygdala (Kiloh et al. 1974; Mpakopoulou et al. 2008). And Chinese and Russian surgeons have reported success in treating drug addiction through ablation of the nucleus accumbens and cingulate gyrus, again arguably restoring drug addicts’ capacity for self-control (Lekovic 2008).⁶ In a recent article that faithfully captures the core ideas behind this therapeutic and reformatory approach, David Eagleman (2011) writes:

Neuroscience can offer customised, brain-based approaches to rehabilitation, in which people are helped to overcome mental illness, drug addiction or even poor impulse control. This could replace the blunt assumption that prison is always the best approach. ... It is time to let go of our intuitions about how people should behave and pay attention to how they do behave — to run our legal system as rigorously as a science experiment. Careful attention to detail will allow us to clean up the streets with less cost, slow the revolving doors of the penal system and divert resources into effective programmes rather than simply build more prisons. A brain-based approach can be more cost-effective, humane and successful. If we desire our medical treatments to be biologically informed, shouldn’t we demand the same from our courtrooms?

And echoing similar sentiments Reinhard Merkel and colleagues conclude that:

Should a brain-invasive treatment for severe psychopathy ever become available, subject to common criteria for weighing the potential risks [and] benefits... nothing stands in the way of offering such a treatment to people in preventive detention if that is the only alternative to their being kept in custody indefinitely. (Merkel, Boer et al. 2007:382)⁷

The senses of “responsibility” related to therapy and reform differ from those mentioned in the previous sub-section. Here the idea is that by restoring people’s mental capacities we might either make them into *fully responsible* moral agents or *more responsible* in the sense opposite to “irresponsible” (Vincent 2009). But again we encounter a familiar idea—namely, that responsibility could at least in principle be *restored* through the restoration of mental capacities.

⁴ Although Khalifa and colleagues also find “evidence that phenytoin (a drug used to treat epilepsy) could help to reduce the intensity of impulsive aggressive acts in people with antisocial personality disorder” (Khalifa et al. 2010), they stop short of recommending pharmacological interventions to treat antisocial personality disorder since, in their view, too few studies have been conducted to warrant such a recommendation.

⁵ In these trials therapy is only a supplement rather than a replacement for justice since inmates they receive no concession on their sentences for participation.

⁶ Objections to these procedures are mentioned in section “[Problems with the Restoration of Responsibility](#)” below.

⁷ Michael Gazzaniga (2011:57–58) is also enthusiastic about possible future treatments for psychopathy, and such ideas are also explored though without endorsement by Walter Glannon (2011:87–88, 119) and Sean Spence (2008).

Problems with the Restoration of Responsibility

The previous section listed a range of existing and proposed medical treatments for use in various legal contexts, as well as considerations cited in discussions of this topic, which either presuppose or explicitly support the idea that responsibility also tracks restored mental capacities, or, put another way, that responsibility can at least in principle be restored through the restoration of mental capacities. However, this section will consider some objections which, I believe, challenge this extension of capacitarian thinking—i.e. objections which challenge the idea that as long as the right mental capacities are restored then responsibility too will be restored. But before proceeding, let me note two points.

First, the objections which I will highlight usually arise in legal cases where mental capacities are *forcibly* (as opposed to voluntarily) reinstated through *direct brain interventions* (as opposed to through conventional psychotherapy or just waiting for the person to heal unaided by medicine). Although I will not comment on why these objections tend to arise in the context of cases with these particular features until section “[What Is So Special About These Scenarios?](#)” below, it is worth noting that this is why the examples in the previous section all involved mental capacity restoration through direct brain intervention based techniques,⁸ and why many of them also involved forced treatment—i.e. because that is the context in which the objections that I wish to highlight tend to arise.

The second point is that many objections only highlight reasons why we *should not* (at least allegedly) use such interventions.⁹ For instance, some people doubt that these drugs really work—i.e. that they do anything more than just mask the symptoms of insanity, in effect only protecting our sensibilities rather than securing justice. Commenting on the court’s decision to forcibly treat Charles Laverne Singleton with anti-psychotic drugs in order to make him competent for execution, Judge Heaney cites the following reasons for his dissent:

Receiving treatment is not synonymous with being cured. Antipsychotic drugs merely calm and mask the psychotic symptoms which usually return to debilitate the patient when the medication is discontinued... A subset of psychotropic medications, psychoactive drugs diminish the symptoms of mental illness, but they do not cure the underlying mental illness... Singleton is not cured; his insanity is merely muted, at times, by the powerful drugs he is forced to take. Underneath this mask of stability, he remains insane. (*Singleton v. Norris* 2003, Judge Heaney’s dissenting comments, internal quotes and citations omitted)

Others doubt that we can accurately assess whether, when and to what degree such interventions work. For instance, Lindsay Horstman argues:

⁸ It is difficult to tie down the precise meaning of the locution “direct brain interventions”, or to characterize what distinguishes them from *indirect* interventions. After all, presumably any intervention that changes a person’s mental states—e.g. this very sentence, which aims to convey something to the reader—accomplishes this by effecting some change in their brain, and this might be thought to reveal a flaw with this distinction (e.g. see Levy 2007:71, reporting not endorsing). In what follows I will employ Jan Christoph Bublitz and Reinhard Merkel’s suggestion that while “indirect interventions are inputs [via our senses] into the cognitive machinery our minds are adapted to process, ... direct interventions change the cognitive machinery itself” (this issue). Although this way of capturing the distinction still strikes difficulties in hard cases—e.g. see Neil Levy’s “Ruby and Elly” example (2007:106–107)—it neatly categorizes the interventions that I discuss here as *direct* brain interventions.

⁹ The list of objections that follows is only intended to be demonstrative not exhaustive.

Drugs inherently have only a temporary effect on people, usually just as long as the drug is in the blood stream [and] the temporary nature of chemical competence renders difficult a reliable assessment of the sufficiency of an inmate's competence for execution. (Horstman 2001–2:846–7, internal citations and quotation marks omitted)

Moral objections also abound. Leah Eisenberg (2004) worries that it might be unethical—or at least that it might violate codes of medical ethics—for the medical profession who are meant to save lives to instead treat people to make them competent for execution. Lene Bomann-Larsen's (2011) discussion shows that there are even concerns about criminal offenders “participat[ing] in voluntary [let alone forced] rehabilitation programs aim[ed] at correcting undesirable behaviour” through direct brain interventions. Anjan Chatterjee (2007:305) points out concerns about the risks and side-effects associated with all treatments that involve direct brain interventions, which suggests that unless the benefits are expected to be sufficiently great and are likely to obtain while the risks are sufficiently low and are unlikely to eventuate, then the use of such interventions might be difficult to justify. Furthermore, Wayne Hall points out that psychosurgery—for instance, the previously-cited Russian and Chinese treatments for drug addiction through ablation of the nucleus accumbens and cingulate gyrus and the surgical removal or lesioning of the amygdala—are highly controversial practices with a disturbing track record (2006). And Hank Greely elaborates that although “success in eliminating heroin addiction [was reported by the Russian and Chinese groups,] it is not clear what else was eliminated” (2009:179).

Others highlight legal and political objections. For instance, Stacey Tovino (2007:49–50) and Bublitz and Merkel (this issue) maintain that people have a right to cognitive or mental liberty—or, as Martha Farah puts it, “the freedom to think one's own thoughts and have one's own personality” (2002:1126)—which might be infringed by such interventions. And Clive Staples Lewis argues that the therapeutic approach would allow the state to inflict unspeakable brutality onto citizens whose values it disapproves of:

[I]f crime and disease are to be regarded as the same thing, it follows that any state of mind which our masters choose to call ‘disease’ can be treated as crime; and compulsorily cured. [But] one school of psychology already regards religion as a neurosis[, and w]hen this particular neurosis becomes inconvenient to government, what is to hinder government from proceeding to ‘cure’ it? [W]hen the command is given, every prominent Christian in the land may vanish overnight into Institutions for the Treatment of the Ideologically Unsound. [T]he Humanitarian theory... carries on its front a semblance of mercy which is wholly false. (Lewis 1963:229)

The above objections raise *countervailing reasons* against using such interventions. However, the objections that I shall consider next arguably challenge the very idea that as long as the right mental capacities are restored then responsibility too will be restored—i.e. they challenge the claim that responsibility *would* even be restored rather than merely raising doubts about whether we *ought to* (attempt to) do this.

Worries About Medical Interventions and Justice

Consider yet again the defendant's objection to being medicated in *Riggins v. Nevada* (1992). Looking at it one way—probably the conventional way—his objection was that this would deprive him of a fair trial, because the drugs would make it hard for the judge and jury to observe first hand, and thus to properly appreciate, the degree of his insanity. Viewed from this angle, although the defendant's point was valid—in fact it even led the

court to overturn the forced medication order—it was only a procedural legal objection. The defendant was entitled to ensure that the judge and jury had access to all of the relevant evidence—in this case, evidence from first hand observation of his composure in court while he was un-medicated.

However, viewed from another angle, what the defendant was arguing was that if he were put on trial in a medicated state, then the person who would stand before the court would in a *qualitative* sense be a very different sort of person to the one who committed the crime. The person who committed the crime was an insane individual and their actions were motivated by disordered thinking and by insane reasons, whereas the person who would have stood before the court if the forced medication order had not been overturned would have been a sane individual, and the kinds of explanations that the latter could give for the former’s actions could not accurately convey (whether through their explicit utterances or connotatively) the disordered nature of the thinking and the insanity which led to the commission of that crime. What the court needed to hear in order to accurately adjudicate responsibility was not a sanitised version of an insane person’s disordered reasoning—hearing that might only have inculpated the defendant by presenting an un-representatively pasteurised version of his story—but rather the irrational, disturbed and disturbing thinking of the person who committed that crime.

When *Riggins* is viewed from this angle, the objection against forced medication reflects not only a simple procedural matter directed at ensuring a fair trial. Rather, it reflects the deeper worry that in order to ensure that the person who stands trial is sufficiently similar to the one who committed the crime, we may have to resist the urge to make an incompetent person competent, since the medications may not merely affect their ability to give testimony but also its content. Similar considerations also apply to the case of Jared Lee Loughner who is charged with killing six people including Chief US District Court Judge John Roll and leaving another fourteen injured including US Representative Gabrielle Giffords in Tucson, Arizona, on January 8, 2011 (e.g. see Perry 2011; or Emshwiller and Audi 2011). Although it appears that the untreated Loughner is not competent to stand trial, the problem with medicating him to make him competent—assuming, that is, that the treatments would even work—is that doing this would put a very different person (yet again, I mean different in a qualitative rather than a numerical sense) on trial to the one who originally committed that crime.

But does this “different people” objection apply only to cases like *Riggins* and Loughner in which the defendant was insane *when they committed the crime*, or does it also apply to cases in which they only became insane *after committing the crime*? Cases of the latter sort present a special challenge because while in the former cases the person was arguably never even responsible for what they did and so it makes less sense to restore their competence in order to put them on trial or to punish them, in the latter cases the person was originally responsible for what they did and so it might make more sense to restore their competence to do these things.¹⁰ I believe that at least sometimes the “different people” objection might apply to the latter kinds of cases, and the following example will help me to explain why:

There are cases ... in which DRT [dopamine replacement therapy, used in the treatment of symptoms of Parkinson’s disease,] induces behaviour that individuals claim [is] authentic. For example, one male who became fascinated with anal sex following DRT claimed that he had these desires prior to DRT treatment but was too

¹⁰ My point is not that there is no reason whatsoever to attempt to restore the competence of the former kind of person in order to put them on trial, but rather that even if we succeed in restoring their competence they ought not be expected to answer *for their crime* but at most only to answer *to the charge* (see Duff 2007:24).

embarrassed to act on them. The medication allowed him to ‘realise these desires’. His interest in these sexual behaviours stopped following a change in his medication, and he later expressed regret at his behaviour. (Carter, Ambermoon et al. 2011:95–6)

Carter and colleagues claim that this example presents a problem for responsibility attribution, and I agree that it does. But the reason why I think that this example is especially pertinent to my argument is because the problems for responsibility arise not due to concerns about the person’s *competence*—after all, the person remained competent during the DRT treatment—but due to concerns about the effects that some medications might have on a person’s *authenticity* or *personal identity*.¹¹ The claim is, in effect, that some medications might make (or leave) a person competent, but nevertheless *too different* to be treated as the same person.

Klaming and Haselager (2010) make a similar point, though their discussion relates to the way in which deep brain stimulation (DBS) affects personal identity and thus responsibility. After briefly reviewing the literature on dissociative identity disorder (DID) and responsibility, and comparing the discontinuities and changes in character and personality experienced by people with DID to those experienced by DBS patients, Klaming and Haselager suggest that:

It is difficult to determine responsibility for action if ... disruptions in psychological continuity lead to wrongful or even criminal behavior. Based on the literature on the responsibility of DID patients, it can be argued that if... similarity of character of a patient [is] disrupted as a result of the DBS treatment, it is not intuitively clear that he should be held (completely) responsible for his actions. (2010:8)

Similar worries about the effects of DBS, informed by past experience and empirical case studies, lead Felicitas Kraemer to recommend “that we must do further research on the relevance of authenticity and alienation to patients treated with DBS” (2011:1). And Inneke Bolt and Maartje Schermer’s discussion of pharmacological interventions used in the treatment of attention deficit and hyperactivity disorder (ADHD), as well as to putatively enhance cognition, leads them to conclude “that the question about identity and psychopharmacology is not merely an academic issue, but one that influences everyday lives of real people” (2009:103). In short, mounting empirical evidence substantiates the worry that direct brain interventions might have adverse effects on such things as authenticity and personal identity by significantly altering character and personality.

Admittedly though, while Klaming and Haselager (2010:7) cite some Dutch cases which suggest that courts sometimes accept that changes in personality or character can diminish and even extinguish responsibility, as I already hinted at earlier this is by no means how courts always or even usually approach such cases. Often courts view such cases as presenting the following challenge: how to disentangle the behaviour produced by the drugs (i.e. the drugs’ side effects, or simply their failure to adequately treat the disorder) from behaviour that can be rightly attributed to that person. For instance, in the case of Martin (2011) the defendant, Terrence Lewis Martin a former Minister of Parliament in the Australian state of Tasmania, was charged and “found guilty by a jury of one count of sexual intercourse with a young person [a 12 year old girl], and one count of producing child exploitation material [photographs of the same 12 year old girl]”. However, although his 10 month prison sentence was suspended, the rationale provided for this was not that he

¹¹ Contrast this to Kennett and Matthews (2002) who argue that DID diminishes responsibility by affecting mental capacities required for agency.

underwent changes in his character or personality—things that might have been relevant to authenticity and personal identity—but because of “overwhelming evidence that Mr Martin’s hypersexuality[, which is what lead to his behaviour, was] directly attributable to the dopamine agonist for treatment of his Parkinson’s disease”—i.e. causal responsibility was attributed not to *him* or even to a *changed/different him*, but to something completely different and separate from him, namely the DRT medications.

When such cases are approached from this more conventional angle, what we have is essentially an epistemic problem—i.e. “Was it the drugs or was it the person?”. But what rationale is there for supposing that this question even *has* an answer—i.e. that it is possible (and I mean possible, not just whether we know how to do it, or even whether we have the observational or computational resources to do it) to disentangle behaviour produced by the drugs from the person’s own behaviour? Put another way, what reason is there to suppose that the problem with which we are dealing is epistemic rather than ontological? Carter and colleagues’ as well as Klaming and Haselager’s examples suggest that even subjects themselves can be dumbfounded by their own predicament—that even they may not know whether to attribute behaviour to themselves or to the medications that they are on. But if even the subjects themselves can not answer this question despite their epistemic privilege—i.e. despite their incorrigibility with regards to knowledge of this domain—then what reason is there to suppose that this question even has an answer, other than a lingering nagging suspicion that perhaps they are just trying to pull the wool over our eyes in order to avoid punishment?

In light of this worry, the most neutral way to view such cases is that what we have post-treatment is not a *person* and their *medications* but just a *treated person*, albeit one that from a qualitative perspective might be significantly different to their pre-treatment self. We should thus not discount the possibility that in at least some cases—i.e. at least when some interventions are used—this post-treatment person may not be sufficiently similar to the pre-treatment person to warrant putting the post-treatment person on trial to assess the degree of the pre-treatment person’s responsibility for what was done, or to expect them to take responsibility for it, or to hold the post-treatment person responsible for what the pre-treatment person did. My point is that even if interventions used to restore competence for trial or for punishment succeed in making a person competent, if their likes, dislikes, preferences and so on—loosely, their personality or character—are sufficiently changed, then *they* may no longer be the person who should be tried or punished. It might not be legitimate to expect a treated person to answer for their earlier self’s actions because the content of their testimony might be unrepresentative due to their altered character, or to attribute responsibility and to hold responsible the treated person for what their earlier self did, because we might now be dealing with someone who is in no position to answer (usually to society, but perhaps even to themselves) for the actions of the person that originally committed the crime.

Whether this objection will apply to the scenarios with which this paper concerns itself depends on whether the interventions used to restore competence for trial and punishment might sometimes have similarly profound effects on people’s authenticity and personal identity to effects that disorders like DID and interventions like DRT and DBS sometimes have. And admittedly, I have said nothing to show that this is the case. However given what is at stake (i.e. a potential miscarriage of justice) and that it is generally not uncommon for psycho-active drugs to have effects on such things as mood and temperament that contribute to character and personality, the burden of proof surely lies with those who wish to use such interventions to make people competent for trial and punishment to show that they are unlikely to have sufficiently profound effects in this regard.

Writing about the “medicate to execute” debate, Barry Latzer (2003:9–10) argues:

...the ‘artificiality’ of the sanity make[s no] difference. If a psychotic were to commit a crime while on medication – medication that made him aware and in control of his conduct – the artificiality of his sanity would provide no defense. Just as culpability for the commission of a crime depends on the cognitive and volitional state of the actor at the time, and not the cause of the actor’s condition, the competence of the death row inmate should not be affected by the basis for that competence. (2003:9)

However, if I am right about the significance of the mounting empirical evidence about the effects that some direct brain interventions might have on authenticity and personal identity via their effects on character and personality, then regardless of whether competence (for standing trial or for punishment) would be restored, what Latzer’s analysis overlooks is that the reason why it might be unjust to put on trial or to punish a post-treatment person for a pre-treatment person’s actions is that the post-treatment person may not be sufficiently similar to the pre-treatment person to warrant doing these things. If what we want to achieve by putting people on trial and by punishing them has anything to do with responsibility—e.g. with assessing the degree of a person’s responsibility for what *they* did, expecting them to take responsibility for what *they* did, and holding *them* responsible for it—then mere competence will not suffice because competence in the sense of having the right mental capacities is not the only factor that matters when it comes to responsibility. Rather, it also matters that the person whose responsibility will be assessed, who will be expected to take responsibility for what was done, and who will be held responsible for it, is sufficiently similar to the one who committed the crime, and this is something that might be undermined by certain ways of restoring mental capacity.

Worries About Using Medical Treatments for Therapy and Reform

My concerns about using medical treatments for therapy and reform within the criminal law context are different to the concerns outlined above, because the senses of “responsibility” involved in the context of therapy and reform differ from those involved in the context of justice. Here, the putative aim of mental capacity restoration¹² is to make people into *fully responsible* moral agents (i.e. to ensure that they have what I call “capacity responsibility”), or to make them *more responsible* in the sense that is opposite to them being “irresponsible” (i.e. to increase the level of what I call their “virtue responsibility”) (Vincent 2011b:18 and 16 respectively).

However, as I will shortly argue, precisely because of what is involved in being *fully responsible* and *more responsible*, the aim of *making* people responsible in these senses may simply be unattainable. But first let me explain what I think is involved in “being responsible” in these two senses. Consider Jane and John:

Jane is 8 years old and very well behaved. She gets up in the morning all by herself, she washes and gets dressed and even makes her own lunch, she doesn’t fight with other kids at school, after school she does her homework, she cleans up after herself, she helps me make dinner, and she even looks out for her older brother John. However, despite the fact that Jane is such a responsible little girl, I also know that she is ultimately only a little girl – she still lacks the mental capacities that one needs to have

¹² I refer to this aim as *putative* because, as I explain at the end of this sub-section, I worry about the distinction between *mental capacities* and *character flaws*.

in order to be a fully responsible person – and so I do not really blame her when she sometimes fails to do these things. Jane is a responsible person, even though she is not yet a responsible person; or put in a less ambiguous way, Jane is a responsible little girl, even though she is not yet a fully responsible person.

Now consider John, Jane’s older brother. John is 17 years old and not at all like his much younger sister — some would call John an irresponsible young man. He wags school, he won’t make his own lunch even though bought lunch is much more expensive, if and when he gets to school he always gets in some kind of trouble with the teachers, his bedroom is a pigsty and we are always cleaning up after him in the rest of the house; he can not even be trusted to look after his little sister. Nevertheless, despite the fact that John is an irresponsible young man, we all know him well enough to know that he is actually a fully responsible person – we’ve seen what he is capable of doing when he puts his mind to things – and that is precisely why we are so very dark on him for his constant misbehaviour. John is not a responsible person, even though he is a responsible person; or put in a less ambiguous way, John is an irresponsible young man, even though he is in fact a fully responsible person.

[A]s regards their *capacity* responsibility, Jane is not yet a fully responsible person but John is a fully responsible person[; and] as regards their *virtue* responsibility, Jane is a responsible little girl and John is an irresponsible young man. (Vincent 2009:115-6)

Sometimes we make assessments about whether someone is responsible in the *virtue* sense—i.e. we assess their *character*—and when we say that they are responsible (or not) in this sense, what we mean is that they can be admired (or condemned) on account of who they are. But at other times we make assessments about whether someone is responsible in the *capacity* sense—i.e. we assess those of their mental capacities that are relevant to moral agency—and when we claim that they are responsible in this sense (or not), what we mean is that they are (or are not) legitimate candidates for attributions of praise and blame on account of what they do. However, if this is an accurate account of these two *senses* of “responsible” and of these senses’ respective *significance* in regards to different kinds of appraisal, then it is not clear that people could indeed be *made* responsible.

On the first point, fully responsible moral agents are meant to be legitimate targets of praise and blame for what they do. However, as John Fischer and Mark Ravizza point out, “a moderately reasons-responsive mechanism could be ‘implanted’ by a scientist, or produced via direct ... stimulation of the brain” (1998:230), and “an agent need not be held morally responsible for acting on ... moderately reasons-responsive mechanisms that have been ‘implanted’: these mechanisms are *not the agent’s own*” (1998:235, emphasis added). Their point is that when one person meddles with another’s mind, the former (the “manipulator”) becomes at least partially responsible for what the latter one (the “manipulated subject”) does for at least some period of time afterwards. This is certainly why responsibility is undermined in their manipulation examples, and the leading intuitions there regarding mental capacity *ownership* are precisely what I think creates the problem for the aim of *making* people responsible in the capacity responsibility sense. Thus my initial worry is that once we meddle with a person’s mind to restore their capacity responsibility, paradoxically they may cease to be legitimate targets of praise and blame for what they do.¹³

¹³ A similar argument is also endorsed by Bublitz and Merkel’s earlier piece (2009). Shaw (this issue) maintains that this argument is not convincing if its intention is to establish that manipulations diminish responsibility by diminishing free will. Shaw’s critique does not pose a problem for my argument though because free will does not play a role in what I say.

Admittedly, on Fischer and Ravizza's account this situation would only obtain until the manipulated subject makes their new mental economy *their own*—until they have “taken responsibility” for their implanted mental capacities. However, despite Fischer and Ravizza's optimism about whether it is possible for a manipulated person to genuinely take responsibility for their new manipulated self in retrospect, and precisely how they might do this,¹⁴ I am a lot more pessimistic. Part of my pessimism stems from the fact that if people are treated non-voluntarily or involuntarily—and the therapeutic context is all about people whose competence is presumptively compromised (if it weren't compromised, then we shouldn't be giving them therapy), so this will of necessity entail at least a diminished kind of voluntariness—then the resulting capacities that they might acquire will not be *their own* and thus their possession of those capacities will not confer “fully responsible” status onto them anyway.¹⁵ In time this may change—they *may* “grow into” their new selves—but we might still have non-insignificant grounds to worry that direct brain interventions which implement large-scale changes in one fell swoop could sever psychological continuity between the pre-treatment and the post-treatment person, and that this in turn would give us grounds for concern that the post-treatment person is not in an authoritative position to retrospectively endorse those changes. After all, the person who would be conducting the retrospective assessment might be too different from the person who they were prior to the treatment, and so any endorsement that the post-treatment person gives might be insufficiently *partial* in respect of the pre-treatment person to warrant the claim that a legitimate endorsement of those changes could be made in retrospect.¹⁶

On the second point, a person who is responsible in the sense that is opposite to being “irresponsible” is meant to be a legitimate target of admiration on account of who they are. However, when we reflect on what kind of reactive attitude would be appropriate to someone who is *made* responsible, it is not clear that such a person would be a legitimate target for such reactive attitudes. In the novel “A Clockwork Orange” Anthony Burgess writes the following about Alex, the hyper-violent young protagonist who semi-voluntarily undergoes a procedure intended to make him into a responsible individual in the sense of being a legitimate target for admiration on account of who he is:

¹⁴ In Chapter 8 of *Responsibility and Control*, John Fischer and Mark Ravizza outline their views about “the process by which a mechanism leading ... to an action, becomes *one's own*” (1998:207, original emphasis). They argue that there are certain processes through which a person can *retrospectively* come to own the mental capacities which they have come to possess, and they call these processes “taking responsibility” (I normally use this expression to mean something different namely to accept blame or liability, or to answer to a charge). However, although I find most of their compatibilist theory to be very compelling, for reasons outlined briefly in this paragraph I do not endorse what they say about this matter. Fischer and Ravizza have surprisingly little to say about this kind of scenario—here's what they say: “[T]he agent's view of himself must be based on his evidence in an appropriate way. Obviously, this is abstract and schematic. This condition is intended (in part) to imply that an individual who has been ... induced to have the relevant view of himself ... has not formed his view of himself in the appropriate way. But the relevant notion of appropriateness must remain unanalyzed” (1998:236). There is admittedly debate about whether the tension between the structural and historical components of Fischer and Ravizza's theory is real or only apparent (e.g. see Stump 2002; or Shabo 2005), however for the reasons briefly outlined here I take this to be a less satisfactory element of their theory.

¹⁵ We may even doubt whether genuine consent can be given in “It's either prison or treatment for you!” style conditions (Bomann-Larsen 2011).

¹⁶ Shaw (this issue) adds that Fischer and Ravizza's suggestion that just making the manipulated person aware of the fact that their views have been implanted through direct brain interventions would make them responsible for what they henceforth do, because it would now be up to them to reflect and to endorse (or not) these views as they see fit, is implausible. For the reasons that I have outlined here, I agree with Shaw.

Does God want woodness or the choice of goodness? Is a man who chooses the bad perhaps in some way better than a man who has the good imposed upon him? (Burgess 2000:71)

Burgess' point is that reform which is inflicted upon one person by another from the outside has little if any value—i.e. that moral virtue of this sort is not something that can be implanted into or inflicted onto people. If Burgess's intuitions are veridical, and I suspect that they are, then the very aim of *making* people responsible in this sense of the word would be flawed at a very fundamental level, because no amount of other-inflicted changes could ever *make* one into a legitimate subject of admiration on account of who one now is—i.e. into a responsible person in this virtue sense. At best, one might become a well-behaved puppet or automaton, or perhaps a prisoner of the “do gooder” impulses that have been implanted in one's psyche like ever-watchful and ever-knowing security guards, but puppets and automata who were once something more are simply not subjects of moral admiration or condemnation. Rather, they are the proper subjects of our pity, since their choices express the values that someone else has implanted in them.¹⁷

What Is So Special About These Scenarios?

Before concluding, let me summarize the previous section's main claims and explain why the objections that I highlighted tend to arise in contexts where mental capacity is restored through direct brain intervention based techniques.

The previous section listed four objections to using direct brain intervention based mental capacity restoration techniques for the pursuit of the criminal law's aims of justice, therapy and reform. The first two relate to justice. First, I argued that if a defendant was insane when they originally committed a crime, then it makes little sense to restore their competence in order to assess their responsibility, because this would put a very different person (in a qualitative sense) on trial to the one who committed the crime. Second, I argued that if a defendant became insane only after committing a crime, then it might still be objectionable to restore their competence if the intervention used to do this would sufficiently alter their character or personality, since this too would result in a very different person being tried or punished to the one who committed the crime. Third, in relation to therapy, because responsibility transfers from manipulated subjects to their manipulators, and a manipulated subject might be insufficiently partial to their pre-manipulated self to be able to reliably endorse their manipulation in retrospect, I therefore argued that by meddling with defendants' minds we might, rather paradoxically, rob them of the opportunity to be legitimate subjects of praise and blame for what they do. And fourth, in relation to reform, I argued that pity not admiration is the reactive attitude that would be fitting to someone who was made responsible in the sense that is opposite to “irresponsible”.

To some extent it is an empirical fact, demonstrated by the mounting empirical evidence, that some direct brain interventions have significant effects on subjects' authenticity and personal identity. Empirical studies, not philosophical inquiry, are needed to explain why this occurs and the degree to which different medical interventions have such effects.

¹⁷ My point is not that authenticity diminishes simply because the change occurred due to a direct brain intervention—an idea that Shaw (this issue) considers and rejects—but because the manipulator's values override those of their manipulated subject. Also note that Shaw's discussion at that point in her article relates to praise/blame on account of what one does, whereas my point here relates to admiration/condemnation on account of who one is.

However, it is also partly a conceptual matter, because a number of the features which make such techniques attractive paradoxically also bring about the related worries, so let me first highlight these attractions. First, to the extent that direct brain interventions aim to target the organic causes¹⁸ of mental incapacities by mending or supplementing dysfunctional brain mechanisms,¹⁹ treatments which employ such techniques hold out the promise of greater *effectiveness* (both in terms of the *degree* of change and of its *permanence*) which hopefully prevents recidivism, as well as greater *speed of action* which is particularly attractive given the high cost of keeping prisoners incarcerated (see Shaw, this issue). Second, only *minimal cooperation* is needed from subjects of direct brain intervention based treatments, since (e.g.) once the right pharmaceutical is running through a subject's veins, mental capacity restoration will take effect regardless of their efforts. This is likely to make such techniques particularly attractive in cases where a subject is so incompetent that they lack the ability to take part in more indirect techniques like cognitive behavioural therapy (CBT) which require one to play an active role in one's own treatment, because their inability to play that role will not jeopardize the outcome. And third, relatedly, this also makes such techniques into perfect candidates for *forced* and *coerced* treatment, since there is relatively little that the subject can do to resist their effects, and this feature is especially attractive in the context of pursuing justice where subjects might perhaps otherwise prefer to remain incompetent but (e.g.) stay out of the line of fire of justice—e.g. to stay alive rather than being made competent only then to be tried and subsequently executed for their crimes.

However, these same features are also the source of the philosophical objections. A technique which is *too effective* may change a person too radically, and one that acts *too quickly* may sever the link between their former and their latter self. Both of these features have the undesirable effect—undesirable, that is, from the perspective of what matters for responsibility—that the post-treatment person may understandably feel that they have little in common with the pre-treatment person, that they can neither understand nor endorse their earlier decisions and actions, and hence that they should therefore be liable neither to have to answer nor to be punished for their earlier self's crimes (as previously noted, there might still be reason to expect them to answer *to the charge* even if they need not answer *for their actions*). It is a genuine conundrum that the potential *effectiveness* of direct brain intervention based techniques—the *degree* of change that they might accomplish, those changes' *permanence*, and the *speed* with which those changes take effect—could be a source of concern, since it would presumably be counter-intuitive to suggest that another technique would be preferable on account of its being *ineffective*. Nevertheless, it seems that in this context when it comes to effectiveness there can be too much of a good thing. And concerns about authenticity and manipulation also arise due to the *lack of input* into how one will eventually be transformed through such techniques—the most that a person undergoing such treatments can do is to give a nod of approval for the intervention to be made, but we may have little appreciation for or input into how we will be transformed—and the fact that such techniques can even be used on someone *against their will* just adds to the difficulties encountered later on by manipulated people who might be too alienated from their earlier selves to legitimately retrospectively approve the outcome of the procedure, even if they wanted to do this.

Naturally, problems remain. In what follows I shall quickly mention four such problems and then offer brief responses to indicate how the above discussion might help us to

¹⁸ Neil Levy addresses worries about interventions that only treat symptoms (2007:121–123).

¹⁹ This especially applies where knowledge of the brain-based causes of a mental dysfunction is used to develop treatments that target those causes.

address them. Firstly, it is not as if we normally have that much control about how our mental makeup changes anyway when indirect techniques such as CBT are used instead of direct brain interventions. Rather, something changes on the inside and all we know is that we now feel better and that we identify with the new self. But since we do not suppose that this poses a problem for responsibility, why should we suppose that the techniques discussed in this paper would pose problems? A tentative reply to this objection might be that CBT achieves its changes slowly and in small steps—these are two components of the *effectiveness* of direct brain interventions which, as I argued above, might plausibly make a difference when it comes to responsibility—and hence that there is a morally significant difference between these two ways of changing people’s minds after all. Secondly, it might also be objected that it’s not as if we ever gave consent to starting out with the particular psychological makeup that we were given at birth, so why should it make a difference when something similar happens without our consent later on in life? A reply here might be that adults have developed personalities, and that this imposes a justificatory burden on anyone who wishes to intervene to alter them. A third objection might be that people can undergo profound conversion experiences while serving a long prison sentence, but since we do not release them early on account of this, why should things be any different when it comes to similarly profound changes brought about by direct brain interventions—i.e. why should people who are significantly altered through direct brain interventions escape being held responsible for what they have done? One reply here might be to question the propriety of the current practice of keeping people in prison even after they have undergone such profound changes—i.e. to argue that the current practice of sentencing people to life without even the chance of parole regardless of how much they change does not necessarily provide a legitimate norm—or to point to the fact that in prison such changes usually happen slowly, which entails (given what I said above) that their authenticity and personal identity are not necessarily affected, and thus that they (unlike the subjects of direct brain interventions) can be kept in prison for the duration of their sentence. Fourthly, and finally, is it not conceivable that some people might be quite comfortable with fast-paced changes which leave them radically altered, and thus mightn’t it be reasonable to use medical treatments on such people to promote the aims of the criminal law? My tentative reply here is that unless we have some way to determine whether a given person falls into this category or not, then we should give everyone the benefit of the doubt to avoid inflicting injustice onto those whose identity would be significantly altered.

The above list of problems and my brief responses are only demonstrative—I list them to explain why I believe that more work needs to be done to fully address the issues discussed in this paper, but also to provide a rough indication of the sort of work that I think needs to be done.

Concluding Remarks

This paper’s main point can be stated as follows: what might suffice for the restoration of legal competence need not necessarily suffice for the restoration of responsibility. Irrespective of whether having the right mental capacities is all that it takes for a person to be legally competent or not, for responsibility it also matters *how* those mental capacities came about—i.e. via which *aetiology*—because some aetiologies might fail to preserve additional factors of importance to responsibility such as authenticity, personal identity and mental capacity ownership.

My concerns about the effects of direct brain interventions on authenticity, personal identity and mental capacity ownership sprang from two separate sources. First, I cited mounting empirical evidence which suggests that at least some direct brain interventions have non-insignificant effects on their subjects' authenticity and personal identity by significantly changing their character and personality. Consequently, I argued that in light of what is at stake, the burden of proof lies with those who wish to use such techniques to restore mental capacity in the legal context, to establish that it is sufficiently unlikely that the proposed techniques would have such effects. Thus, echoing Felicitas Kraemer's (2011) recent suggestion, it would be helpful to conduct further empirical research into the effects that direct brain intervention based mental capacity restoration techniques might have on personality and character in order to ascertain their possible effects on authenticity, personal identity, and mental capacity ownership. Second, I also argued that the very same features which according to David Eagleman (cited towards the end of section "Capacitarianism and Restored Mental Capacities" above) make direct brain interventions so attractive within the context of the criminal law—i.e. their *speed* of action, their touted *effectiveness*, and the fact that *little cooperation* or even *input* is needed from the person being treated for the interventions to take effect—are paradoxically also what gives rise to the cited concerns about authenticity, personal identity and mental capacity ownership.

Thus, although responsibility might indeed track mental capacity, we cannot assume that responsibility would necessarily also track restored mental capacities, or, put the other way, that mental capacity restoration would necessarily restore responsibility. This conclusion is important for at least two reasons.

First, the capacitarian idea that responsibility tracks mental capacity is very prominent in the neurolaw literature. So prominent in fact, that what might once have been viewed as a *character flaw* and thus as a legitimate ground for condemnation or at least as an aggravating consideration at sentencing, may these days often be viewed as a *mental incapacity* and may thus be treated as if it were a potential excuse or mitigating factor. The literature on psychopathy provides an excellent example (e.g. Kröber and Lau 2000; Maibom 2008; Reimer 2008). On one account psychopaths have significant deficits in affective mental capacities, and if such mental capacities are required for fully responsible moral agency then psychopaths may need to be at least partially excused for what they do, and (the retributive component of) their sentences may need to be mitigated. That was the capacitarian take on psychopathy. However, there is also another plausible description of psychopaths, one that is framed not in terms of what capacities they have or lack (i.e. affective mental incapacities) but rather in terms of who they are (i.e. nasty individuals), and this latter way of describing psychopaths (we might call it "characterian" for contrast with the capacitarian description) hardly seems like a reason for excuse or mitigation. The tension between these competing descriptions of psychopathy can be captured using several related dichotomies:

mental incapacity	—	character flaw
cannot	—	does not
mad	—	bad
disease	—	person
illness	—	depravity
disorder	—	difference
has	—	is

Do psychopaths have mental incapacities or character flaws? Is it that they cannot understand what they ought to do or control their actions in light of such an understanding, or is it rather that they just do not exercise their cognitive and volitional capacities? Are psychopaths mad or are they bad? Is their behavior the expression of a mental disease or is it just them doing what they genuinely want to do? Are they mentally ill or simply depraved? Are psychopaths' brains disordered or merely different? Do these individuals have psychopathy, or is it rather that some people are psychopaths?

The problem is not that these dichotomies exist, but that we do not yet have a principled way to choose which description is more appropriate in a given case—i.e. to decide whether (e.g.) to call someone *mad* or to call them *bad*—but yet the moral, legal and medical ramifications of these labels differ radically—i.e. excuse, mitigation and treatment for the mad; condemnation, aggravation and punishment for the bad. And yet such dichotomies permeate (media coverage of) legal debates about prominent criminal cases. For instance, Howard Greenberg (2011), attorney for Levi Aron who is accused of murdering 8 year old Leibi Kletzky, exploits precisely this family of dichotomies when he comments on his intention to pursue an insanity defense for his client, who reportedly sustained a serious head injury as a child, and whose family has a history of schizophrenia. Greenberg explains:

It's a simple case. It's a perfect storm when you combine traumatic brain injury with schizophrenia. He's either evil or he's crazy, and we intend to prove that he is crazy.

The same dichotomy has also played a central role in the media's coverage of the legal case surrounding Anders Breivik's massacre of 77 people in Norway on July 22, 2011. To this day the question that is repeatedly asked is whether Breivik is bad or mad, evil or insane—i.e. whether his actions expressed his evil character or his mental incapacities, and whether he should be treated with anti-psychotic drugs or condemned for who he is and blamed as well as punished for what he did (e.g. France-Presse and Agneze 2011; The Telegraph 2012). Finally, here is how the Australian media reported on the awful case of 37 year old Martin Freeman who on January 29, 2009 flung his own little daughter over the side of the Westgate Bridge as his other children and numerous stunned motorists looked on in horror:

Defence lawyer David Brustman called on the jury to consider why a man would commit a "truly horrible" crime. "Very few cases could induce more prejudice ... There in the dock sits a man who flung a four-year-old girl, his own daughter, to her death," he said. "Now, how *bad* does that get? Is this simply the face of pure *evil*?" Mr Brustman argued the killing of the girl was done by a *mentally impaired* man rather than a father intent on punishing his ex-wife in the worst possible way. (Farnsworth 2011)

As I argue elsewhere (Vincent 2011a), the capacitarian approach favours interpretations of legal evidence that employ concepts from the left hand side of the above list of dichotomies. Evidence is almost inevitably taken to support the claim that defendants are mad in the sense that they have mental incapacities, illnesses, diseases, or disorders, in virtue of which they could not have done what it is claimed that they ought to have done, and so excuse and mitigation is allegedly called for. However, since an exclusively capacitarian approach leaves no room for concepts on the right hand side of the above list of dichotomies, and yet these concepts also seem to have a place in our everyday legal and moral discourse, it is therefore a useful exercise to highlight ways in which capacitarianism might have overlooked important parts of the moral landscape—in the case of the present

paper, to highlight exceptions to the capacitarian idea that responsibility tracks mental capacity.

Second, there is sometimes an air of nonchalance in judicial comments about medically treating defendants or inmates. For instance, at a recent workshop in Australia on the topic of neurolaw held at the Supreme Court of New South Wales (Vincent, Hall & Kennett 2011) I was surprised that apart from concerns about effectiveness and possible adverse health effects, it appeared that none of the Judges, Queen's Counsel or prosecutors could see the use of mental capacity altering direct brain interventions in anything other than an unequivocally positive light. The concerns which I voice in this paper—namely, that the aims of justice, reform and therapy might be jeopardised rather than promoted by restoring people's competence through the use of direct brain interventions—fell largely on deaf ears, which is why I have again tried to highlight these concerns here.

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

References

- Aristotle. (1976). Book three: Moral responsibility—Two virtues. In J. Barnes & H. Tredennick (Eds.), *The ethics of aristotle* (pp. 111–141). London, England: Penguin Books.
- Atkins v. Virginia. (2002). 536 U.S. 304 (USA).
- Bolt, I., & Schermer, M. (2009). Psychopharmaceutical enhancers: Enhancing Identity? *Neuroethics*, 2(2), 103–111.
- Bomann-Larsen, L. (2011). Voluntary rehabilitation? On neurotechnological behavioural treatment, valid consent and (In)appropriate offers. *Neuroethics* (OnlineFirst).
- Bradford, J. M. W., & Pawlak, A. (1993). Double-blind placebo crossover study of cyproterone acetate in the treatment of the paraphilias. *Archives of Sexual Behavior*, 22(5), 383–402.
- Bublitz, J. C., & Merkel, R. (2009). Autonomy and authenticity of chanced personality traits. *Bioethics*, 23(6), 360–374.
- Burgess, A. (2000). *A clockwork orange (with an introduction by Blake Morrison)*. London, UK: Penguin Books Ltd.
- Carter, A., Ambermoon, P., et al. (2011). Drug-induced impulse control disorders: A prospectus for neuroethical analysis. *Neuroethics*, 4(2), 91–102.
- Chatterjee, A. (2007). The promise and predicament of cosmetic neurology. In W. Glannon (Ed.), *Defining right and wrong in brain science* (pp. 302–311). New York: Dana Press.
- Curtis. (1999). *Connecticut v. Kenneth Curtis*.
- Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, Massachusetts: MIT Press.
- Duff, R. A. (2007). *Answering for crime: Responsibility and liability in the criminal law*. Oxford, UK: Hart Publishing.
- Eagleman, D. (2011) The human brain: turning our minds to the law. Accessed Jul 01, 2011 from <http://www.telegraph.co.uk/science/8428793/The-human-brain-turning-our-minds-to-the-law.html>.
- Eisenberg, L. (2004). Medicating death row inmates so they qualify for execution. *Virtual Mentor* 6(9). Accessed May 11, 2011 from <http://virtualmentor.ama-assn.org/2004/09/hlaw1-0409.html>.
- Emshwiller, J. R., & Audi, T. (2011) Federal court sides with loughner in drug debate. *Wall Street Journal Online*, July 12, 2011. Accessed Aug 09, 2011 from <http://online.wsj.com/article/SB10001424052702303678704576442641078371886.html>.
- Farah, M. J. (2002). Emerging ethical issues in neuroscience. *Nature Neuroscience*, 5(11), 1123–1129.
- Farnsworth, S. (2011) Dad guilty of Westgate Bridge murder. Australian Broadcasting Corporation, story dated Tuesday March 29, 2011. Accessed April 05, 2011 from <http://www.abc.net.au/news/stories/2011/03/28/3175976.htm>.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge, UK: CUP.

- France-Presse, A. (2011). Norway split on Breivik's likely fate in mental ward, as mass-killer himself 'insulted' by ruling. National Post, November 30, 2011. Online at <http://news.nationalpost.com/2011/11/30/norway-split-on-breiviks-likely-fate-in-mental-ward-as-mass-killer-himself-insulted-by-ruling/> Accessed Apr 05, 2011.
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy*, 66(3), 829–839.
- Gazzaniga, M. S. (2011). Neuroscience in the courtroom. *Scientific American*, 304(April), 54–59.
- Glannon, W. (2002). *The mental basis of responsibility*. UK: Ashgate Publishing Limited.
- Glannon, W. (2011). *Brain, body, and mind: Neuroethics with a human face*. New York, NY, USA: Oxford University Press.
- Greely, H. T. (2009). Who knows what evil lurks in the hearts of men?: Behavioral genomics, neuroscience, criminal law, and the search for hidden knowledge. In N. A. Farahany (Ed.), *The impact of behavioral sciences on criminal law* (pp. 161–179). New York, NY: OUP.
- Greenberg, H. (2011). Transcribed portion of video interview available at http://www.youtube.com/watch?feature=player_embedded&v=1W974a8dr7o Original story dated 22nd December 2011 at http://failedmessiah.typepad.com/failed_messiahcom/2011/12/video-levi-arons-attorney-explains-inbreeding-remark-567.html Both URLs Accessed Jan 04, 2012.
- Hall, W. D. (2006). Stereotactic neurosurgical treatment of addiction: Minimizing the chances of another 'great and desperate cure'. *Addiction*, 101(1), 1–3.
- Hart, H. L. A. (1968). IX. Postscript: Responsibility and retribution. *Punishment and responsibility* (pp. 210–237). Oxford, UK: Clarendon Press.
- Horstman, L. A. (2001–2002). Commuting death sentences of the insane: A solution for a better, more compassionate society. *University of San Francisco Law Review* 36, 823–852.
- Kennett, J., & Matthews, S. (2002). Identity, control and responsibility: The case of dissociative identity disorder. *Philosophical Psychology*, 15(4), 509–526.
- Khalifa, N., et al. (2010). Pharmacological interventions for antisocial personality disorder. *Cochrane Database of Systemic Reviews (Online)*(9).
- Kiloh, L. G., Gye, R. S., et al. (1974). Stereotactic amygdalotomy for aggressive behaviour. *Journal of Neurology, Neurosurgery and Psychiatry*, 37, 437–444.
- Klaming, L., & Haselager, P. (2010). Did my brain implant make me do it? Questions raised by DBS regarding psychological continuity, responsibility for action and mental competence. *Neuroethics*, OnlineFirst: 13 pages.
- Kraemer, F. (2011). Me, myself and my brain implant: Deep brain stimulation raises questions for personal authenticity and alienation. *Neuroethics* OnlineFirst: 15 pages.
- Kröber, H. L., & Lau, S. (2000). Bad or mad? Personality disorders and legal responsibility—The German situation. *Behavioral Sciences and the Law*, 18, 679–690.
- Latzer, B. (2003). Between madness and death: the medicate-to-execute controversy. *Criminal Justice Ethics*, 22(2), 3–14.
- Lekovic, G. P. (2008). Neuroscience and the law. *Surgical Neurology*, 69, 99–101.
- Levy, N. (2007). *Neuroethics: Challenges for the 21st century*. Cambridge, UK: Cambridge University Press.
- Lewis, C. S. (1963). The humanitarian theory of punishment. *Res Judicatae*, 6, 224–230.
- Macey, J. (2010). Antidepressants may help violent offenders. ABC News Online.
- Maibom, H. L. (2008). The mad, the bad, and the psychopath. *Neuroethics*, 1(3), 167–184.
- Martin. (2010). Tasmania v. Martin No. 2 2011 TASSC 36 (Austl.).
- Martin. (2011). Tasmania v. Martin 2010 TASSC 51 (Austl.).
- Merkel, R., Boer, G., et al. (2007). *Intervening in the brain: Changing psyche and society*. Berlin, DE: Springer.
- Mpakopoulou, M., Gatos, H., et al. (2008). Stereotactic amygdalotomy in the management of severe aggressive behavioral disorders. *Neurosurgical Focus*, 25(1), E6.
- Riggins v. Nevada 1992 504 U.S. 127 (USA).
- Perry, T. (2011) Suspected Tucson gunman can be forced to take antipsychotic drugs, judge rules. *Los Angeles Times*, June 30, 2011. Accessed July 01, 2011 from <http://www.latimes.com/news/nation/world/nation/la-na-loughner-20110630,0,6329506.story>.
- Reimer, M. (2008). Psychopathy without (the Language of) disorder. *Neuroethics*, 1(3), 185–198.
- Robinson, P. H., & Dubber, M. D. (2007). The American model penal code: A brief overview. *New Criminal Law Review: An International and Interdisciplinary Journal*, 10(3), 319–341.
- Sapolsky, R. M. (2004). The frontal cortex and the criminal justice system. *Philosophical Transactions of the Royal Society of London*, 359, 1787–1796.

- Bublitz, J. C., & Merkel, R. (this issue). Crimes against minds: Mental manipulations, harms and a human right to mental self-determination. *Criminal Law and Philosophy*.
- Ford v. Wainwright 198 477 U.S. 399 (USA).
- Sell v. U.S. 2003 539 U.S. 166 (USA).
- Singleton v. Norris 2003 319 F.3d 1018 (USA) Accessed on Feb 06, 2012 from <http://caselaw.findlaw.com/us-8th-circuit/1213175.html>.
- Shabo, S. (2005). Fischer and Ravizza on history and ownership. *Philosophical Explorations*, 8(2), 103–114.
- Sher, G. (2009). The searchlight view. *Who knew? Responsibility without awareness* (pp. 3–22). New York, NY: OUP.
- Spence, S. A. (2008). Can pharmacology help enhance human morality? *The British Journal of Psychiatry*, 193, 179–180.
- Stump, E. (2002). Control and causal determinism. In S. Buss & L. Overton (Eds.), *Contours of agency: Essays on themes from Harry Frankfurt*. Cambridge, MA: MIT Press.
- Tancredi, L. R. (2005). The bad and the mad. *Hardwired behavior: What neuroscience reveals about morality*. New York, NY: CUP.
- Shaw, E. (this issue). Direct brain interventions and responsibility enhancement. *Criminal Law and Philosophy*.
- The Telegraph (author not shown). (2012). Norway killer Anders Behring Breivik 'is not psychotic'. The Telegraph, January 7, 2012. Online story at <http://www.telegraph.co.uk/news/worldnews/europe/norway/8993195/Norway-killer-Anders-Behring-Breivik-is-not-psychotic.html>. Accessed Jan 08, 2012.
- Tovino, S. A. (2007). Functional Neuroimaging and the Law: Trends and directions for future scholarship. *The American Journal of Bioethics*, 7(9), 44–56.
- van de Poel, I. (2011). The relation between forward-looking and backward-looking responsibility. In N. Vincent, I. v. d. Poel, & J. v. d. Hoven (Eds.), *Moral responsibility: Beyond free will and determinism* (pp. 37–52). New York: Springer.
- Vincent, N. A. (2009). Responsibility: Distinguishing virtue from capacity. *Polish Journal of Philosophy*, 3(1), 111–126.
- Vincent, N. A. (2010). On the relevance of neuroscience to criminal responsibility. *Criminal Law and Philosophy*, 4(1), 77–98.
- Vincent, N. A. (2011a). Madness, badness and neuroimaging-based responsibility assessments. In M. Freeman (Ed.), *Law and neuroscience, current legal issues* (Vol. 13, pp. 79–95). UK: Oxford University Press (OUP).
- Vincent, N. A. (2011b). A structured taxonomy of responsibility concepts. In N. A. Vincent, I. Van de Poel, & J. van den Hoven (Eds.), *Moral responsibility: Beyond free will and determinism* (pp. 15–35). The Netherlands: Springer.
- Vincent, N. A., Hall, W., & Kennett, J. (2011). *Report on "NeuroLaw in Australia" workshop*, hosted at Macquarie University, Sydney, Australia, on July 14–15, 2011. Published online by the Academy of the Social Sciences in Australia. Accessed May 11, 2012 from <http://www.assa.edu.au/programs/workshop/workshop.php?id=85>.
- Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.