Singapore Journal of Legal Studies
[2025] 35–56

THE FUNDAMENTALS OF THE INSANITY DEFENCE

Michael **Grainger***

This paper engages with fundamental questions about the categorisation of the insanity defence in relation to other defences and about the moral and legal significance of satisfying its requirements. Firstly, it defends Simester's account of insanity as an exemption from moral assessment against Marcia Baron's argument that insanity is really an excuse. Secondly, it argues that we must distinguish conceptually between cases of insanity based on involuntariness, and cases of insanity where the defendant's conduct was not involuntary. It tentatively suggests that certain defendants who, under Singapore's revised insanity defence, successfully plead that they have no "power to control" their actions, should be treated as failing to satisfy the criminal law's underlying voluntariness requirement.

I. Introduction

The insanity defence does not always operate in the same way. Sometimes, it involves a denial of *mens rea*. Other times, the defendant's conduct is involuntary or automatic. On yet further occasions, there is a simple denial of moral agency (as when D acts intentionally but cannot comprehend the moral significance of his conduct). This is well known. However, a proper understanding of why insanity applies so multifariously to exculpate defendants requires a proper analysis of the fundamental principles underlying the general part of the criminal law. This is what Andrew Simester gives us in his recent book, *Fundamentals of Criminal Law*[1]. Simester's subtle account of the complex interaction of these principles and their relationship to doctrines of the general part provokes questions about how insanity should be categorised in relation to other defences, and about the moral and legal significance of satisfying its requirements. This paper takes up two such issues. Firstly, it defends Simester's account of insanity as an *exemption* from moral assessment against Marcia Baron's argument that insanity is really an excuse.[2] Secondly, the paper argues that we must distinguish conceptually between cases of insanity based on involuntariness, and those based on irresponsibility falling short of

---

* Fellow, Tembusu College, National University of Singapore. I want to thank the students who enrolled in my Tembusu Senior Learning Experience Project for helpful discussion of the issues addressed in this paper.

1 A.P. Simester, *Fundamentals of Criminal Law: Responsibility, Culpability and Wrongdoing* (Oxford: Oxford University Press, 2021) [Simester, *Fundamentals*].

2 Marcia Baron, "Excuses and Exemptions: Is it Really a Mistake to Understand the Category of Excuses to Include Infancy and Insanity?" (2023) Criminal Law and Philosophy <doi:10.1007/s11572-023-09698-0> [Baron, "Excuses and Exemptions"].

involuntariness. It tentatively suggests that certain mentally unwell defendants who, under Singapore's revised insanity defence,[3] successfully plead that they have no "power to control" their actions, should be treated as failing to satisfy the criminal law's underlying voluntariness requirement.[4]

II. The Taxonomy of Defences

Insanity is just one of many criminal law defences known to common law jurisdictions, something that raises the question of where insanity "fits in" within the organisational structure of criminal law defences, and of how it differs from other defences such as duress. In *Fundamentals*, Simester has given an extended argument for his preferred taxonomy of defences, arguing that the insanity defence sits within a category of "irresponsibility" defences that "exempt" the defendant from blame by denying his moral agency.[5] Such defences are distinct from "excuses", which merely reduce or defray blame for a defendant who nonetheless acts as a moral agent. In order to assess the import and persuasiveness of Simester's argument, as well as Baron's challenge to it, it is instructive first to look in this section at how defences have generally been organised by criminal law scholars, before turning to examine Simester's position more closely in sections III and IV.

There are, of course, many ways for defendants to argue that they are not culpable for committing a "*pro tanto* criminal offence" (that is, for satisfying the *mens rea* and *actus reus* elements of an offence). In addition to insanity, pleas such as self-defence, necessity, duress and infancy are available as defences to the commission of a *pro tanto* offence in all common law jurisdictions. The successful invocation of any of these defences entails that the defendant was not sufficiently culpable to deserve a criminal conviction. Yet the fact that each defence *shares* this culpability-denying feature should not also be taken as a denial that each has its own unique moral basis, and thus its own internal logic which reflects that moral basis. There is, for example, no point in requiring that a defendant pleading necessity for his use of force must be acting to repel a threat from another human being. That requirement is only appropriate in the *self-defence* context, which has a different moral basis. In this, we can discern a purpose to enacting distinct defences into law, rather than having a unitary, broad and shapeless defence of "non-culpable commission of a *pro tanto* offence". By marking more specific moral distinctions, we help give form

---

3 Penal Code 1871 (2020 Rev Ed) [Penal Code 1871 (S'pore)], s 84.

4 A requirement on which the Singapore Penal Code is strangely silent: Chan Wing Cheong, "Non-Insane Automatism and the Singapore Penal Code: *Public Prosecutor v Ong Jun Yong* [2022] SGMC 37" (2023) 35 SAcLJ 395 [Chan, "Non-Insane Automatism"]. Aligning with other Commonwealth jurisdictions, the existence of such a requirement in Singapore is thought to be implicit by leading scholars. Stanley's Yeo's work on insanity and the voluntariness requirement is particularly useful here: Stanley Yeo, "Mental Impairment" in Chan Wing Cheong, Stanley Yeo & Michael Hor, *Criminal Law for the 21st Century – A Model Code for Singapore* (Singapore: Academy Publishing, 2013) at 58; Stanley Yeo, Neil Morgan & Chan Wing Cheong, *Criminal Law in Singapore* (Singapore: LexisNexis, 2022); Stanley Yeo, "Fleshing Out Malaysian Perspectives on Automatism: *Abdul Razak bin Dalek v Public Prosecutor*" [2011] 1 SJLS 289; Stanley Yeo, "Putting Voluntariness Back Into Automatism" (2001) 32(2) VUWLR 387; Stanley Yeo, "The Insanity Defence in the Criminal Laws of the Commonwealth of Nations" [2008] 2 Sing JLS 241.

5 Simester, *Fundamentals*, *supra* note 1, see *eg* 15, and generally.

to the law. We make it easier for legal officials to apply it and for the rest of us to follow it.

A preliminary point. Defences like self-defence, necessity, duress and infancy fall with the broad category of *substantive* defences. In labelling them as "substantive", I mean to distinguish them from "procedural" defences, such as diplomatic immunity or unfitness to plead. The latter pleas are available to both culpable and morally innocent defendants alike, so we can assume nothing about the moral blameworthiness of the defendant who successfully invokes one of those defences. By contrast, the substantive defences allow us to conclude that the defendant lacked sufficient culpability to warrant conviction, even if the route to that lack of culpability varies across the various defences.

One famous distinction *within* the substantive defences is between those defences that *justify* the commission of the *pro tanto* offence, and those which merely *excuse* it.[6] Justifications and excuses both allow us to conclude that the defendant was not culpable for his *pro tanto* offence. Indeed, this is implied by the ordinary meaning of the words "justified" and "excused". Despite this, the reasons that a justified defendant lacks culpability are not the same as the reasons for which a defendant is excused. Justified defendants are non-culpable because they acted *permissibly*. That is to say, it was, morally and legally, okay on this occasion to act as they did. There is nothing for them to be culpable *for*, no breach of duty for the criminal law to disapprove. Excused defendants, on the other hand, have acted impermissibly.[7] However, there is some *additional* reason not to blame them, some factor about them and/or their conduct that alters our response to their impermissible conduct.

Some examples of the justification/excuse distinction may help to illustrate this point. John, defending himself from an unprovoked onslaught from Mark, commits no crime if he intentionally uses reasonable force that harms Mark.[8] John had a right to act as he did. His plea is therefore of justification, and self-defence is the paradigmatic justification defence. Contrast John with Matt, who strikes Victor at the behest of Luke. Luke has threatened to kill Matt unless he assaults Victor. In these circumstances, Matt's natural plea is of duress, which in this kind of situation is merely excusatory because Victor is an innocent third party with the right not to be

---

6 This distinction is admitted to exist in some form by virtually all criminal law scholars, even if the contours and importance of the distinction are controversial. For some important recent contributions, see: Victor Tadros, *Criminal Responsibility* (Oxford: Oxford University Press, 2007); R A Duff, *Answering for Crime: Responsibility and Liability in the Criminal Law* (London: Bloomsbury, 2007) [Duff, *Answering for Crime*]; John Gardner, *Offences and Defences: Selected Essays in the Philosophy of the Criminal Law* (Oxford: Oxford University Press, 2007) [Gardner, *Offences and Defences*]; Douglas Husak, "On the Supposed Priority of Justification to Excuse" (2005) 24(6) Law & Phil 557. All of these writers are indebted to the discussion in George P Fletcher, *Rethinking Criminal Law* (Oxford: Oxford University Press, 2000).

7 It is not an entirely individualised assessment. (This distinguishes excuses from accounts that say it is "purely" about the actor: Michael S Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford: Clarendon Press, 1997) at 482.) Rather, while excuses start with the actor's reasons, they subject those reasons to objective review. Duress is an excuse only because any of us in a similar situation might have done the same.

8 Like most theorists, Simester also happens to believe that the self-defending defendant commits no legal wrong. However, unlike some, Simester does not think the absence of wrong is definitional of justifications, because he thinks that some cases of necessitous actions concede the commission of a wrong. The claim is interesting but for present purposes incidental.

assaulted. Matt gets a defence not because he had a *right* to act as he did, but because he demonstrated no less fortitude than the average person in the face of such a serious threat. It is his understandable failure to do what was righteous that grounds his defence. We have said that all excuses must have an "additional" culpability-denying factor, one that explains why we withhold blame from the defendant despite his impermissible conduct. It is the "understandability" of the defendant's conduct in the face of a threat that plays this culpability-denying role in the particular context of duress.

At this point, it might be asked why we need to make this distinction between justifications and excuses. It has already been observed that each substantive defence doctrine rests on its own unique moral basis, and that its moral basis has implications for the proper structure of each such legal doctrine. Why do substantive defences need to be allocated to a further sub-class of defences, in addition to being members of the class of substantive defences? The answer is that there are certain practical payoffs to the justification/excuse distinction. For one thing, understanding which defendants are justified and which are merely excused has implications for who can assist and/or resist the use of force by the defendant. Self-defenders like John ought to be helped by any police officer who observes the conflict. Mark had no right to attack John and is morally liable to suffer justified force not only from John, but also from anyone else who knows the facts of the situation and wants to assist John. Matt's situation is quite different, in that a police officer, forced to choose between Matt and Victor, ought to protect Victor, even at the cost of harming Matt. While we do not blame Matt for his conduct, Victor still deserves the protection of the state and his fellow citizens. He is entirely innocent.[9]

III. Simester's Taxonomy: Insanity and Irresponsibility

On the account of excuses offered so far, we can see that insanity, too, would sit on the excuse side of the line. If Peter attacks Anthony because of a psychotic delusion that Anthony has committed a terrible wrong and that he (Peter) is God's chosen instrument of divine vengeance, there can be no serious suggestion that Anthony loses his right not to be attacked for this reason. Just like Matt, Peter has no entitlement to use force against Anthony. Again, just like Matt, we withhold our blame from Peter due to the circumstances in which he commits what remains an impermissible act. We have different reasons for excusing Peter than we do for excusing Matt, to be sure. There is no suggestion that we are sympathetic to Peter on the basis that he understandably failed to exhibit sufficient moral fortitude in response to a threat. But Peter's illness is a relevant factor that prevents us from attributing culpability for an impermissible action. Moreover, the practical implications of the justification/excuse distinction work the same for both Peter and Matt. We would also want the police officer to intervene to protect Anthony at the cost of harming

[9] What Simester shows more clearly than most others, though, is that the payoff of this emphasis on lawfulness is that State actors can be justified, whereas they cannot be excused. The State does not get the indulgence that we show to human frailties when excusing someone. See Simester, *Fundamentals*, *supra* note 1 at 416.

Peter. At first glance, therefore, there seems to be every reason to categorise Peter's conduct as excused.

This conclusion, however, rests on the proposition that justification and excuses are exhaustive of the moral categories of substantive defences. Many writers take exactly this view. They believe that insanity is properly categorised alongside duress and infancy in a "grab-bag" of fully excusatory pleas for impermissible action. For Simester, however, this minimalist approach is unhelpful. He suggests that there is a third important category of substantive defences – "irresponsibility defences" – that sits alongside justifications and excuses. Along with infancy, a plea of insanity denies the culpability of the defendant in a way profoundly different from a defence like duress. For Simester, an insane (or infant) defendant is outside the realm of moral agency for the purposes of the criminal law, such that he is not an appropriate candidate for moral evaluation at all: he is "morally non-responsible". A defendant successfully pleading duress, by contrast, is excused because of the *quality* of his moral reasoning. His failure to exercise sufficient moral fortitude is understandable because he has responded to a threat in a manner that any of us might have done. An insane defendant's reasoning is not like that. In an important sense, it is irrelevant how that reasoning measures up to the choices a normal person would have made. Just as there is no need to "excuse" a tiger for eating its keeper in a freak accident, there is no need to excuse Peter if he was not the kind of agent who can be subject to culpability evaluations in the first place.

Interestingly, Simester is so strongly committed to his additional category that he claims the distinction between irresponsibility defences and all other substantive defences is significantly more *important* – more "basic" – even than the well-known distinction between justification and excuse.[10] For Simester, there is a crucial respect in which justifications and excuses like duress are fundamentally similar – that they both speak to the defendant's reasons for acting, *ie*, his rationale. Admittedly, they do so in different ways. In terms of the philosophy of practical reasoning, justifications identify an undefeated reason on which the defendant acts, whilst duress identifies a defeated reason that fails to ground culpability.[11] Nonetheless, both defences are "rationale-based" and thus gain their culpability-denying force from the engagement of a competent actor with the reasons for and against acting as he does. ("'Yes, I did it,' she says, '*and here's why*.'"[12]) Irresponsibility defences, on the other hand, exempt the defendant from blame. They do this by showing that the defendant was not (sufficiently) morally responsible for his conduct, in the sense that he was not (sufficiently) responsive to moral reasons. We cannot and should not assess whether his conduct was reasonable or understandable. Those terms are reserved for evaluating the reasoning of competent moral agents.

For Simester, therefore, the picture is as follows:

1. Justifications deny impermissibility;
2. Excuses deny that a moral agent was to blame in the circumstances; and
3. Irresponsibility defences deny that the actor had sufficient moral agency in the first place.

---

10 Simester, *Fundamentals*, *supra* note 1 at 423.

11 A view also taken by Gardner, *Offences and Defences*, *supra* note 6 at 109.

12 Simester, *Fundamentals*, *supra* note 1 at 422.

And, just as there are practical payoffs to being cognisant of the justification/excuse distinction, Simester suggests that the acceptance of his trichotomous picture brings important insights. On my reading, the payoff of the irresponsibility defences/rationale-based defences distinction is that it helps us to see where requirements of reasonableness ought and ought not to find their way into legal doctrine. Irresponsibility defences do not, and should not, contain any requirement of reasonableness. But, absent that distinction, it may be tempting for courts to "read in" requirements that insane defendants act "reasonably" in some sense in order to qualify for an "excuse".

There are at least two junctures at which such a misstep might occur. The first is in the development of doctrine grounded in the famous passage in the *M'Naghten* rules which provides an insanity defence where the defendant laboured under "[a] partial delusion only, and is not in other respects insane, . . . [she] must be considered in the same situation as to responsibility if the facts with respect to which the delusion exists were real."[13] The interpretation of this rule is unclear and inconsistent across jurisdictions, but the strangeness that has sometimes been remarked of trying to "enter into" the defendant's delusions and assess their reasonableness within that world is a recognised problem. The failure to recognise the category of irresponsibility defences perhaps goes some way to explaining why that might be so. Another example, given by Simester himself, is that irresponsibility defences ought to be available to all offences, as there is no "weighing" of countervailing interests to be done – that weighing is characteristic of reasonableness assessments. The difference can be seen in the rule that duress is not available to murder in many jurisdictions. Again, this rule is controversial. Nonetheless, reasonable disagreement about its desirability is possible when we are talking about *excusatory* duress. Simester's category of irresponsibility defences allows us to see that any such rule in the context of irresponsibility defences would be inappropriate: there is no question of the weight that an insane defendant must give to another's interests in his reasoning if the defence is not weighing up the force and quality of his reasoning at all.

IV. Distinguishing Exemptions from Excuses: Baron's Challenge

Simester gives a plausible and detailed account of what distinguishes irresponsibility defences like insanity from excuses like duress. Through unpacking the foundational principles underlying the criminal law, Simester argues persuasively that insanity (and infancy) *exempt* D from the assessment of moral blameworthiness. Excuses, on the other hand, defray blame: they are *part* of – internal to – that culpability assessment. This distinction between exemption from and defrayment of blame tracks two of Simester's foundational principles: respectively, moral responsibility and culpability.

While Simester is not alone in organising defences in this way,[14] Marcia Baron has recently questioned this approach. For Baron, Simester's approach wrongly

---

[13] *Daniel M'Naghten's Case* (1843) 8 ER 718 [*M'Naghten's Case*] at 723.

[14] Duff and Gardner have particularly well-developed arguments for such a taxonomy – see Duff, *Answering for Crime*, and Gardner, *Offences and Defences, supra* note 6.

withholds moral responsibility from too many defendants. Using the example of infancy, she writes:[15]

> Children below the age of criminal responsibility but (well) beyond babyhood are appropriately viewed as moral agents…. We hold them responsible, but at the same time we view them as far less blameworthy for wrongdoing than older children or adults. Their age is an excusatory factor but does not exempt them from moral responsibility.

Baron suggests that we should therefore jettison the idea of irresponsibility defences tracking moral responsibility in favour of an approach that acknowledges that they "serve sometimes to exempt, but only at the limit; in their less extreme form, they excuse. It is thus better to understand them as excuses and to allow that sometimes an excuse can exempt."[16] In other words, we should jettison any fundamental distinction between irresponsibility defences and excuses. Instead, all such defences should be assimilated to excuses, because they are paradigmatically about defraying blame (except "at the limit"). The payoff of her approach is that we – and even the law – can make more fine-grained distinctions of blameworthiness depending on context:[17]

> The impairment could be a temporary condition, as in the case of children, an impairment that is gradually diminishing and is expected to disappear; or a sporadic condition (as in the case of some mental illness); or a condition that is likely to be permanent (as in the case of exceptionally severe mental illness, as well as severe cognitive impairment). In each instance, infancy and insanity can be understood as sometimes full defenses, sometimes partial defenses. In addition, they can be understood as defenses that much of the time would operate not as blanket defenses but instead as defenses to specific actions.

Still, while Baron highlights the potential benefits of a more fine-grained framework, she concedes that "pragmatic" constraints may sometimes demand a coarser approach to moral responsibility in the law than in morality: "It is true that, practically speaking, we have to draw a bright line and specify a minimum age for criminal responsibility. (Whether we really have to do something similar with respect to insanity is a matter for debate.) But that we need to do so does not entail that the conceptual framework has to be tailored accordingly."[18] Even where such practical compromise is necessary, however, Baron sees her approach as having a further payoff: we are better able to see those practical constraints for what they are, rather than pretending that we are tracking a pre-existing moral bright line.

If Baron is correct that Simester's account cannot accurately capture relevant distinctions in culpability, then his account should be revisited. However, I think she

---

15 Baron, "Excuses and Exemptions", *supra* note 2 at para 3(b).
16 *Ibid* at para 2.
17 *Ibid* at para 3(b).
18 *Ibid* at para 3(d).

is mistaken. It is useful to identify exactly where the argument goes wrong. Doing so draws out some of the most important characteristics of the insanity doctrine.

Firstly, it is not at all clear, purely as a matter of intuition, that *any* cases falling under the current highly restrictive definition of insanity are rightly classified as partially or entirely suitable for culpability assessments (i.e. as cases in which the defendant has a "true" excuse). Culpability assessments are, as Simester argues, assessments of rationales. But, when attention is paid to what is required for a defendant successfully to plead insanity, it seems artificial to subject that defendant to any such assessment of his rationale. In England, for example, where the *M'Naghten* rules still constitute the law governing insanity, defendants must meet one of two "cognitive" limbs in order to succeed:[19]

> Every man is to be presumed to be sane, and ... that to establish a defense on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was laboring under such a defect of reason, from disease of mind, and not to *(1) know the nature and quality* of the act he was doing; or *(2) if he did know it, that he did not know he was doing what was wrong*. [emphasis and numerals added]

Most other common law jurisdictions do not far depart from this basic formulation in their penal codes, although some, such as Singapore, have added a further possibility of claiming insanity due to lack of control:[20]

> 84.—(1) Nothing is an offence which is done by a person who, at the time of doing it, by reason of unsoundness of mind, is —
> (a) incapable of knowing the nature of the act;
> (b) incapable of knowing that what he is doing is wrong; or
> (c) completely deprived of any power to control his actions.

It is hard to see how Baron envisages that defendants who are "incapable of knowing the nature of [their] act", "incapable of knowing that what [they are] doing is wrong; or "completely deprived of any power to control [their] actions" could in any sense be suitable candidates for having their reasoning in respect of their actions subjected to gradated evaluations of culpability as moral agents. Indeed, inapt attempts to engage in this kind of investigation by hopeful lawyers are frequently the subject of criticism by clinicians.[21] Baron's focus on the infancy defence may have caused her to miss this point. Insofar as the insanity defence goes, it is unclear whether Baron has established any clear benefit to deleting the category of irresponsibility defences. In other words, it is not clear what fine-grained differences in blameworthiness that insanity reconceptualised as an excuse could capture.

---

[19] *M'Naghten's Case, supra* note 13.

[20] Penal Code 1871 (S'pore). Subsection (c) was added in 2020.

[21] Stephen J Morse, "Causation, Compulsion, and Involuntariness" (1994) 22(2) Journal of the American Academy of Psychiatry and the Law Online 159-180 [Morse, "Causation, Compusion, and Involuntariness"].

Admittedly, Baron is correct that many children caught by the *infancy* defence are moral agents outside of the legal context. At least sometimes, then, there will be cases where a culpability assessment will be appropriate outside of the law, but not accommodated within it. Simester himself admits this, when he says that:[22]

> [A] nine-year-old is at least somewhat of a moral agent. The point of the law's infancy defence, however, is that [they do not have] capacity for moral reasoning *sufficient* to render them eligible for the kind of official blaming judgements that the criminal law involves. [The] child is [not] fully moral responsible; more importantly, [the child] is insufficiently morally responsible.
> [emphasis in original]

Still, even if some infants caught by the defence are moral agents outside of the law, Baron has not established that irresponsibility defences are really "excuses". Again, they contain no requirement of reasonableness or understandability. Indeed, Simester's point that the law requires only "sufficient" moral agency leads us to the second of Baron's apparent mistakes. This is that Baron appears to countenance only a universal, context-insensitive standard of moral responsibility. Simester's account of moral responsibility is best interpreted as positing a context-sensitive standard of responsibility, one that differs between the official context of criminal law judgments and other contexts (such as everyday life). Not only is this interpretation supported by passages like that quoted earlier (referring to the relevant capacity needing to be sufficient for "official blaming judgments"), it seems right that criminal law should be more restrictive in determining who is a suitable object for its judgments. The criminal law is an especially intrusive tool, and maximum latitude should therefore be given before allocating moral responsibility. That point holds as much for infancy as for insanity.

If the foregoing claim is accepted, we should not, *contra* Baron, think of the difference in standards of moral responsibility between criminal law and other contexts as determined merely by pragmatic considerations. Admittedly, there will be an ineluctable element of pragmatism when translating moral responsibility into legal criteria. The underlying capacities relevant to responsibility are scalar properties which will be present to a greater or lesser extent in all actors, and the law must draw a line somewhere. It cannot capture precisely the individualised differences in capacity of each defendant. This inability is, in Baron's terms, a pragmatic consideration that prevents a more subtle tracking of underlying capacities. Similar problems are a familiar feature of debates over where to draw bright-line distinctions such as the age of consent. A specific age is used as a proxy for sufficiency of responsibility, but this approach is necessarily morally over- and under-inclusive given the varying capacities of people of that age.

Nonetheless, such pragmatic considerations speak only to where the line should be crystallised.[23] (If the age of consent is "approximately" 18, should it be exactly 18, or 18 years and 6 months, etc.) This element of pragmatism does not undermine

---

[22] Simester, *Fundamentals*, *supra* note 1 at 206.

[23] Among many others, see A P Simester & Andreas von Hirsch, *Crimes, Harms, and Wrongs: On the Principles of Criminalisation* (London: Bloomsbury, 2011).

my claim that the general standard of moral agency that requires such crystallisation is higher in the law than for informal moral judgments. The criminal law is not a moral system but a punitive social institution. So, there is no reason why its standards must be identical to those in extra-legal morality. At the same time, the criminal law should be responsive to moral norms. We should see it as morally unjust to subject insane defendants meeting a highly restrictive test like that derived from *M'Naghten* to assessments of their culpability, especially when the context of that assessment is the criminal law, even if we were to accept that some of them are moral agents outside of the law. Similarly, infants who are moral agents for various purposes outside of the law should not for that reason be treated as moral agents within it.

This said, some aspects of Simester's theory sit uneasily with my interpretation that he views moral responsibility as context-sensitive in this way. Baron's alternative "universalising" interpretation is not without textual support. The relevant passages appear in Simester's section discussing whether moral responsibility is relational, and if, so relational to what. If moral responsibility is relational, the defendant does not have moral responsibility in general (as a general "status"), but only in relation either (a) to some particular instance of behaviour (responsibility "for" X) or (b) to some person or thing. Support for Baron's universalising interpretation can be found in Simester's denial that moral responsibility is ever relational in the latter sense (b), *ie*, of being responsible "to" someone. On its face, this denial contradicts the claim that the D could have sufficient moral responsibility in some judgment-contexts (such as the criminal law), but not in others (such as moral judgments made within friendships). Is not the "who" (criminal law judge versus friend) doing the work to support my claim that Simesterian moral responsibility can vary in this way? Perhaps not. It may be that Simester can have this context-sensitive moral responsibility without audience-relativity. We might instead see the standard as relative to the consequences of the judgment, as opposed to the identity of the judge. In other words, it may be the very severe consequences that follow a criminal law judgment that militate in favour of the higher standard for culpability assessments than in other contexts. Were there to be another social context where the moral assessment would have a similarly severe outcome, we might be similarly demanding in our standard for moral agency.

This interpretation is bolstered by Simester's comments that "sufficient" moral agency is what is required. Perhaps it is not "full" moral agency that the law requires to justify its culpability assessments, but "enough". Looked at through that lens, we can see that the law is not attempting to replicate the "moral agent/non-moral agent" line at all. While there may be such a line, past which (presumably) all assessments of rationales are fair game, that is not material for the irresponsibility defences (or indeed for exemptions in other contexts). Rather, each context – the law, the family, the school, *etc* – will permit some level of irresponsibility to coexist with eligibility for judgments within that context, whilst stipulating a floor in capacity below which assessments become inappropriate in that domain. This seems to me to be a very plausible view.

So far, we have seen that a credible argument can be made that standards for sufficient moral responsibility vary by context, and that the criminal law's standard is appropriately higher than other contexts. This explains why, at least in some cases

(notably, infancy), we are right to exempt defendants from the criminal law, even if they are sufficiently morally responsible for other purposes. We have also noted, however, that on current insanity doctrine, there are likely to be very few defendants who satisfy the insanity criteria yet who we would also regard as having moral responsibility for their actions outside of the legal context. The case for categorising insanity as an irresponsibility defence is thus especially clear.

Baron, however, has a further argument in favour of her proposed absorption of irresponsibility into excuses. Despite all the foregoing, it is not clear why we see irresponsibility defences as "fundamentally" distinct from other excuses. After all, insanity is different from infancy, and duress is different from mistaken self-defence. Why not accept that all culpability-denying defences for unjustified offending can be brought under the single category, "excuses"? This approach has the virtue of apparent simplicity. And as Baron correctly points out, it would highlight the similarities between the different doctrines, even if it obscured some differences.[24]

One response might be to dismiss this move as semantic. So long as we are aware of the differences between each doctrine, what does it matter whether we organise them into a larger category of "excuses", or subdivide them? However, Simester's approach contains the seeds of a better answer: the sub-categories help to bring out the way his foundational principles interact in the law. Clarity about their interaction can, in turn, explain why certain defences have particular doctrinal features, and why those features should not mistakenly be thought relevant to other defences outside of that sub-category. We have already seen examples of this. The fact that duress is rationale-based, assessing defendants' reasons (and reasonableness) for acting as they did, implies that reasonableness assessments ought to appear in any good definition of that defence. Action taken under duress is not involuntary. It is not even analogous to being involuntary. Neither is it voluntary but delusional, like some defendants who qualify under the cognitive limbs of the traditional insanity defence. And that implies a weighing of interests – it is appropriate to ask whether the defendant's fear, for example, was an understandable response to the threatened harm. That will depend, to some extent, on the severity of the wrong the defendant went on to commit. Similarly, mistaken self-defence shares this rationale-based reasonableness requirement.[25] It is excusatory because it rests on a defendant's mistakenly but reasonably believing he is acting for a justifying reason, such as to prevent an attack. If, as it turns out, there is no such attack, we excuse based on the reasonableness of the defendant's belief. As Simester shows, this kind of assessment is entirely inappropriate in the case of infancy and insanity.

Baron denies this point. She suggests that duress is fundamentally *sui generis*. There is nothing that unites it with other excuses as a coherent category that wouldn't also apply to irresponsibility defences. Her point is that all the excuses are fundamentally dissimilar, except that they deny culpability. However, this part of Baron's account is unconvincing. Baron overlooks the clear similarities between mistaken self-defence and duress as excusing based on the reasonableness of the defendant's

---

[24] Baron, "Excuses and Exemptions", *supra* note 2 at para 3(b).

[25] Or at least it should. In Singapore, it does: Penal Code (S'pore), s 79. Unfortunately, many jurisdictions, including England, require only an honest belief in the necessity and proportionality of force. See, *eg*, *Solomon Beckford v R* [1988] AC 130 (PC, Jamaica).

reasoning, a point of connection which goes beyond the mere fact of denying culpability and merits the identification of a separate category. More worryingly, the example she uses to argue that duress is fundamentally dissimilar from other excuses (and thus that the dissimilarity of insanity from duress does not prevent us from categorising insanity as an excuse) is to compare it to involuntary intoxication. She argues that involuntary intoxication is just as different from duress, with its *sui generis* structure, as the irresponsibility defences, but she believes it would be nonsense to remove involuntary intoxication from the category of excuses. This must mean that removing insanity would be equally odd. The error here is that involuntary intoxication is not usually a defence at all – it normally operates via a denial of *mens rea*. Absence of *mens rea* will, of course, generally foreclose a conviction without the need to establish any substantive defence. While voluntary intoxication is now generally thought to inculpate rather than exculpate, on basis of the defendant's prior fault in becoming intoxicated, involuntary intoxication lacks that prior fault aspect. If his intoxication causes the defendant to be unable to form the relevant *mens rea*, then the defendant can simply be acquitted for failure by the prosecution to prove *mens rea*. In that sense, we should be unsurprised that involuntary intoxication is unlike duress – it is a denial of the *prima facie* offence. Admittedly, in some US jurisdictions,[26] involuntary intoxication can excuse even where there is *mens rea*. However, in these circumstances, the defence works like the insanity defence. It is a matter of temporary irresponsibility, just like many cases of insanity. Again, therefore, there is no surprise that it is different from duress. Indeed, it points in the opposite direction to that suggested by Baron. It helps to show the unity of irresponsibility defences and the usefulness of distinguishing them from excuses for wrongdoing. Where involuntary intoxication is legally recognised as denying the defendant's responsibility for an intended action (because of his intoxicated confusion in forming the intention), it would be inappropriate to build any requirement of reasonableness into the involuntary intoxication test. We are able to conclude this because we know that evaluations of reasoning have no place in irresponsibility defences.

The example therefore militates against Baron's position, in that it shows the unity of irresponsibility defences, and their dissimilarity to excuses. Admittedly, aside from mistaken self-defence, Baron is correct that duress looks pretty lonely as a true, rationale-based complete excuse. But this is a feature, not a bug, of Simester's taxonomy. It shows that in many ways, as a matter of fundamental structure, the irresponsibility defences are the more important category, having more members, and denying culpability in a more profound way than does duress. The scarcity of doctrines like duress emphasises that, once a defendant meets a basic standard of responsibility, the criminal law allows very little room for people to fall short of meeting their criminal law duties. Normally, rationale-based excusatory considerations are a matter for sentencing, where they can interact with limitations of capacity, just as they do in the partial defences.

To sum up this section, Simester's taxonomy places insanity among the irresponsibility defences. This is an important category. It highlights the intrinsic unfairness

---

[26] And elsewhere. See *eg* Criminal Code Act 1899 (Qld), s 28(1).

of subjecting irresponsible defendants to assessments of their reasons for action; and it ought to have the salutary effect of warning against reasonableness requirements creeping into their definition.

At the same time, it is worth noting that the exempting effect of insanity for Simester does not have to be "all or nothing". The standard of sufficient moral responsibility is relative to the context in which the defendant's behaviour is being assessed. It is also, as I have emphasised, relational. It need not, as Baron seems to assume, be applied to the actor in general. While the infancy doctrine is a true "status" defence, applying to all actions so long as the age requirement is met, Simester's account of moral responsibility is subtle enough to support the requirement of a nexus between the defendant's mental illness, its symptoms, and the particular behaviour being assessed.[27] This is exactly the kind of insanity defence that we see in most common law jurisdictions. Simester recognises this: "[I]t seems to me plausible that an agent might lack moral reasoning capacities in respect of some but not all of her actions. That possibility is allowed for in the *M'Naghten* Rules, which regulate insanity under English criminal law, and was endorsed in *Kay*."[28] Of course, whether that is the right doctrinal approach in the case of insanity depends on what is true about the defendant's underlying capacities as a general empirical matter, on whether defendants really can meaningfully respond to moral reasons in some contexts but not others. Simester's account of the "fundamental" principle of moral responsibility as a threshold condition which must be met before any culpability assessment can be undertaken could explain either current doctrine or a status-based approach. That seems like a payoff of the account.

V. Distinguishing between Exemptions: Insanity, Irresponsibility and Involuntariness

In sum, Baron's claim that irresponsibility defences are "almost always" excuses is not correct. Simester's view that such defences mark out defendants as exempt from (ineligible for) moral assessment by the criminal law – as criminally irresponsible – is the more persuasive one.

The categorisation of insanity as an "irresponsibility" defence, however, masks an important further distinction within that category. In the most common type of case, the irresponsibility arises despite the defendant's satisfying the criminal law's underlying voluntariness requirement. A defendant whose offending results from delusional beliefs, such as a gunman who believes he is shooting into a flock of geese instead of a crowd of people, cannot deny that he voluntarily fired the gun. His only hope is to point to his insufficient ability to appreciate the nature of his actions. By contrast, some defendants are irresponsible just because they are involuntary. A classic example is the defendant who "commits" the offence in a state of somnambulism. Under the *M'Naghten* rules, akin to many common-law jurisdictions, a somnambulist whose "disease of the mind" arises from an "internal" cause qualifies

---

[27] *Cf, Public Prosecutor v Wang Zhijian* and another appeal [2014] SGCA 58 at [92]. This point is also emphasised in *Public Prosecutor v Khoo Kwee Hock Leslie* [2019] SGHC 215 at [157].

[28] *R v Kay (Robert)* [2017] 4 WLR 121 (CA, Eng), cited in Simester, *Fundamentals*, *supra* note 1 at 78, fn 9.

for the insanity defence in respect of his unconscious misdeeds. After all, he cannot be said to have known the "nature of his act" or that it "was wrong". He easily satisfies the traditional cognitive test for insanity. Nonetheless, the sleepwalker's innocence is properly founded on a more profound claim than mere attenuated moral capacity. It rests on the idea that he did not "commit" the *actus reus* at all. Even if his bodily movements otherwise met the description in the offence definition, those movements were not voluntary. As such, they do not "belong" to him in the way that the law – criminal or even civil – normally requires for legal responsibility to follow. The somnambulist is, in this respect, in the same moral category as a defendant whose leg moves reflexively in response to a doctor's hammer.

Non-criminal lawyers may doubt that this really can be the legal position, even if they are willing to grant that somnambulists are involuntary for the purposes of moral responsibility. If the law recognises that somnambulists fail the voluntariness requirement, why does it handle such cases under the insanity defence? After all, ordinarily,[29] a defendant against whom no *actus reus* is proved has nothing to answer for and therefore no need to plead a defence. The explanation lies in the fact that the insanity defence has what Hogg and Child call an "inculpatory" aspect.[30] Unusually, defendants are *required* to plead the insanity defence where their failure to satisfy an offence element is caused by a symptom of their mental disorder. This puts them in the position of needing to plead a defence to an offence which they otherwise would not be taken to have committed (hence "inculpatory"). It is well-established that this anomalous rule applies to failures of *mens rea* caused, *eg*, by delusions. Importantly for present purposes, it applies equally to involuntariness. The justification typically given for this rule is that society needs to be protected from dangerous-but-innocent people with mental disorders, and only an insanity verdict and the special dispositional powers it triggers can ensure that. The alternative to an insanity verdict and hospital order – an unqualified acquittal for failure of proof – is too risky.

We can see, therefore, that insanity's categorisation as an "irresponsibility" defence, while valid insofar as it goes, requires further refinement if we are fully to understand the defence. While many, indeed most, insanity defences do not raise questions of voluntariness,[31] others do. Somnambulists, for example, are morally irresponsible only because they are also involuntary at the time of their somnambulism.

To some, this observation may seem unimportant. Precisely because insanity applies equally to voluntary and involuntary insane defendants, we need not worry about the difficult boundaries between these two categories. Either way, the outcome is the same: a special verdict of not guilty by reason of insanity or unsoundness of mind[32] and likely confinement.[33] Perhaps we should be glad of the lightened

---

29 Albeit, in Singapore, this point is not settled: see Chan, "Non-Insane Automatism", *supra* note 4.

30 Claire Hogg & JJ Child, "Not Guilty by Reasons Other than Insanity" in A P Simester (ed), *Modern Criminal Law: Essays in Honour of GR Sullivan* (London: Bloomsbury, 2024) at 155-176.

31 Indeed, a clear majority of successful insanity pleas arise from delusional behaviour that is nonetheless voluntary. A person who delusionally believes he is under attack, for example, clearly acts voluntarily when striking his perceived attacker. Nonetheless, his attenuated capacity for moral reasoning grounds his irresponsibility.

32 Penal Code 1871 (S'pore), s 84.

33 Criminal Procedure Code 2010 (2020 Rev Ed), ss 251–252.

theoretical load. But this incurious approach is surely dangerous – losing sight of the principles that justify a defence is always suspect. We cannot be sure whether they may become relevant in hard cases requiring judicial development of the law.

More urgently, perhaps, a deeper understanding of those principles is also essential if we are to be able to intelligently critique proposals for statutory law reform. And there is no shortage of calls for radical reform of the insanity defence, especially in recent years. What is now a significant constituency of writers argues that the insanity defence should be not just reformed but abolished entirely. It is argued by those abolitionists that the defence is discriminatory, precisely because it denies insane defendants the right to put the prosecution to proof of offence elements like intention and voluntariness. Additionally, and separately, pleading the defence makes defendants liable to confinement on the basis that their mental disorder renders them dangerous. Yet innocent *sane* defendants, whom we might have very good grounds for believing to be dangerous, walk free from an unqualified acquittal. At the more radical end of the abolition debate, writers like Minkowitz have argued that the very idea of a line between those who have "capacity" for the purposes of the law and those who do not is demeaning. There are international human rights instruments that arguably support this view.[34]

Against this backdrop of criticism of the insanity defence, a proper categorisation of irresponsible defendants into voluntary and involuntary subtypes is crucial to improve the quality of discussion. Without an understanding of these subtypes, we are unable to properly assess the merits of abolitionist and other critiques. Why? Let us stipulate that there is a class of schizophrenic defendants qualifying for the insanity defence who are voluntary agents. Their particular illness does not impact them, at least on the relevant occasion, in such a way as to deny their actions' basic voluntariness. In their case, we would *need* the insanity doctrine to avoid their being convicted (especially where an offence is one that imposes strict liability: as we shall see, this point becomes especially important when a defendant "acts" while conscious). At least for them, unlike *involuntary* actors, providing an insanity defence would be an example of opening up an escape hatch from state detention, rather than slamming one marked "involuntariness" (or "automatism") shut. This would weaken the abolitionists' case, because that case is based partly on the unfairness of making defendants take on the burden of pleading an insanity defence, when sane defendants have the opportunity to deny the prima facie case based on mistake, involuntariness *etc*.

Understanding the voluntary/involuntary divide between different cases of insanity is also indispensable to designing the doctrines that might replace the insanity defence if the abolitionists win the argument. Slobogin – an ardent abolitionist – argues that defendants suffering from mental disorders ought to be permitted to plead lack of intention, automatism, *etc*, whenever their mental disorder causes them not to satisfy these elements.[35] Likewise, doctrines of mistaken defence should be amended to allow delusional mistakes to excuse. In his view, we should integrate

---

[34] Convention on the Rights of Persons with Disabilities (opened for signature 30 March 2007), 2515 UNTS 3 (entered into force 3 May 2008).

[35] Christopher Slobogin "An End to Insanity: Recasting the Role of Mental Disability in Criminal Cases" (2000) 86 Va L Rev 1199-1247.

insane defendants into existing doctrines, rather than railroad them into pleading insanity. This cannot be done if we do not know whether involuntariness, lack of intention, or something else (such as a more general "irrationality") underlies the defendant's irresponsibility. Only if he is involuntary should the insane somnambulist be permitted to benefit from exculpation by reason of the automatism doctrine. In sum, this debate about the interaction of irresponsibility defences with doctrines of voluntariness is of far more than academic importance, with implications just as far-reaching as the distinction between irresponsibility and excuse models of the insanity defence.

VI. Conscious Lack of Control: A Test Case

In what remains of this paper, I want to highlight a subset of insanity cases with which, like the courts in many other jurisdictions, the Singapore courts may have to grapple in coming years. These cases lie at the borders of this division between voluntary and involuntary insanity. They are (what I shall call) "conscious lack of control" (CLOC) cases, wherein a defendant seems to lack any power of control over his actions, yet is not in a state of somnambulism or otherwise unconscious action (such as in a state of dissociative state automatism). While I cannot here resolve the debate about their status, I will highlight a very plausible view that these cases can often be examples of involuntary insanity. However, unlike cases of somnambulism, the involuntariness *cannot be explained by the unconsciousness of the defendant*. This requires us to think carefully about the morality and phenomenology of these cases, and to move away from the tendency, of which Simester's *Fundamentals* is arguably an example, to explain the defendant's lack of voluntariness by reference to the defendant's unconsciousness.

The reason that consideration of these cases is of particular interest is because, as we saw earlier,[36] Singapore has recently introduced "lack of control" as a distinct "limb" upon which an insanity defence can be founded.[37] To recap, the law now states:

> Act of person of unsound mind
> 84.—(1) Nothing is an offence which is done by a person who, at the time of doing it, by reason of unsoundness of mind, is —
> (*a*) incapable of knowing the nature of the act;
> (*b*) incapable of knowing that what he is doing is wrong; or
> (*c*) *completely deprived of any power to control his actions*.
> [emphasis added]

---

[36] In Part IV.

[37] As it happens, I have argued elsewhere that we can achieve the right result in the case discussed below in the text by reinterpreting limbs (a) and (b), without requiring limb (c): Michael Grainger, "Insanity and Command Delusions" (2022) 81(3) Cambridge LJ 467. But that is a doctrinal matter. My argument in what follows concerns the more fundamental structure of CLOC cases and their relationship to voluntariness.

Presumably, s 84(1)(*c*) was not designed to cover cases of insane sleepwalking or dissociative state automatism. Admittedly, "unconscious lack of control" (ULOC) cases can be thought of as denying voluntariness and thus as denying the "power to control" one's actions. (Indeed, this is why "sane" somnambulism unproblematically qualifies as automatism, which is widely accepted as representing a denial of the criminal law's underlying voluntariness requirement.) However, it seems unlikely that s 84(1)(*c*) changes anything for defendants pleading those conditions. Even if a sleepwalker's behaviour is being directed by him at a subconscious level, it is unlikely that we will think him able to know "what he is doing is wrong", or even to understand that he is acting at all. As such, limbs (a) and (b) are sufficient for such defendants. This thought is reinforced by the experience of countries that have longer-standing control tests for insanity, such as Australia.[38] In such countries, cases in which a defendant pleading lack of control did not also qualify under the cognitive limbs have been exceedingly rare.[39] Indeed, given this overlap of limb (c) with other limbs of s 84(1), it is unsurprising that there is yet no known use of the limb in Singapore.

It follows from the above that, if limb (c) of the rules is to have any distinct role, it must, at a minimum, apply to defendants in CLOC cases. That is, it must exempt from responsibility defendants who retained a significant degree of consciousness when experiencing a lack of control over their behaviour. Only then is there logical space for the defendant both to understand what he is doing and that it is wrong, but lack the ability to prevent it. In that logical space is where we find the subcategory of cases on which I want to focus, namely those involving delusions of demonic (or deific) possession. Defendants who believe they have been overtaken by a supernatural force often retain a high degree of consciousness. They may well be able to appreciate that what they are doing is not permitted by society or the law. Nonetheless, they may also experience a profound sense of an inability to control themselves.

A. *R v Keal and Demonic Possession*

A good example of an (apparent) demonic possession case was recently heard in the English Court of Appeal. In *Keal*,[40] the defendant was seemingly fully conscious and aware that he was committing a crime. Nonetheless, he felt unable to control his actions due, in his mind, to the work of the Devil. The defendant engaged in an extended violent assault against members of his family, while apologising and explaining the role of the Devil in his actions. As there is no "power to control" limb to the English insanity defence, Keal's defence was denied. The Court reasoned that his apologies indicated that he knew he was doing wrong. However, there is a strong case for saying that Keal was not morally responsible. In Simester's terms, he couldn't "ultimately respond" to the moral reasons applicable to his conduct, even

---

[38] Simon Bronitt & Bernadette McSherry, *Principles of Criminal Law* (Sydney: Law Book Company, 2001) at 221.

[39] A notable exception is *R v Lavell* [2002] WASC 200.

[40] *R v Keal* [2022] 4 WLR 41 (CA, Eng).

though he knew that what he was doing was wrong. (At least, this is a plausible understanding of the underspecified facts in the judgment: for the sake of the discussion here, we can assume that it was the case.)

It seems likely, therefore, that in Singapore, CLOC defendants like Mr Keal would seek to invoke s 84(1)(*c*).[41] Yet, while there may be some degree of shared intuition among readers about Mr Keal's suitability for the insanity defence, there is likely to be much less agreement about the principles that ground his irresponsibility. This is unsurprising. Explaining the exact basis of CLOC defences – and distinguishing deserving cases from undeserving ones – is notoriously difficult.[42] I will not attempt to fully settle that basis here. Our focus is more narrow, and can be summed up in the following question. Let us grant that defendants like Keal are indeed "completely deprived of the power to control [their] actions". Should we further conclude that, akin to our somnambulist, they have also failed to meet the criminal law's voluntariness requirement? I will suggest that the answer may well be "yes", a possibility that may pose problems for the theory of voluntary action articulated in Simester's *Fundamentals*.

Some may think that answer to be obvious, even trivial. If a person is deprived of the power to control his actions, how could that mean anything other than that his behaviour was involuntary? But this is too quick. There are theories claiming to capture the philosophical basis of the "power to control" test which drive conceptual space between that test and the voluntariness requirement. For example, as Stephen Morse reminds us, one explanation (criticised by Morse) of CLOC cases is that the defendant is faced with an unbearable choice between doing an act or suffering some overwhelmingly unpleasant consequence of not doing it.[43] This anticipated consequence might well be real, like the dysphoria of an addict who chooses not to take his preferred intoxicant. Equally, it may be a delusion, like the anticipation of hellfire for disobedience to an imaginary demon. In either case, the defendant is faced with a very constrained choice due to an "internal threat". On that view, his defence is excusatory, analogous in many ways to duress. If we accept the excusatory power of the constraint, arguably we are saying that there is no sufficient power of control over the behaviour. Nonetheless, the act remains voluntary as it is a choice the defendant *could* have resisted.

Like all attempts to cash out the mysterious phenomenon of "lacking control", the unbearable choice theory has some serious weaknesses. It is worth mentioning two. Firstly, and perhaps most troublingly, it lacks clinical evidence. Secondly, even if it does capture the essence of *some* CLOC cases, there is no reason to believe that delusions of demonic possession are always or even often experienced as leaving room for choice, even a difficult one. Very likely, the best interpretation of Keal's situation is that he felt entirely borne along by an alien force controlling his body.

---

[41] Given the dearth of case law (and the fact that the revision dates only from 2019), it is by no means certain that such invocation will be successful. The Penal Code Review Committee Report (2018) does not discuss which specific cases are under consideration, referencing only those with "volitional disorders".

[42] Morse, "Causation, Compulsion, and Involuntariness", *supra* note 21; Stephen J Morse, "Against Control Tests for Criminal Responsibility" in Paul H Robinson, Stephen Garvey & Kimberly Kessler Ferzan (eds.), *Criminal Law Conversations* (Oxford: Oxford University Press, 2009) at 449.

[43] Morse, "Causation, Compulsion, and Involuntariness", *supra* note 21.

In any event, even if there are cases in which the defendant experiences the demonic possession as leaving him with a "difficult-but-not-impossible" choice, we are concerned in s 84(1)(*c*) with cases in which the defendant entirely loses control. More generally, my discussion here is concerned with those cases where the defendant subjectively experiences a total loss of agency. The unbearable choice theory gives us no reason to think that such cases are nonetheless voluntary.

B. *The "Sense of Agency" Analysis*

Disposing of one contrary argument (*ie*, that CLOC cases should be seen as involving voluntary action) does not, however, establish that such cases are *involuntary*. That further conclusion still needs to be justified. There is, I suggest, a preferable analysis that does allow us to make sense of the intuition that defendants like Keal are both morally irresponsible, and that they are not acting voluntarily for the purposes of the criminal law's voluntariness requirement. While I cannot defend this view in full here, I want to highlight that it has considerable explanatory force. The thought is that defendants experiencing CLOC are in a similar situation to the somnambulist discussed above. The somnambulist is widely thought, as Simester puts it, to be "exercis[ing] ... her higher faculties" in that the faculties that make us agents are in play, albeit in her dream state. Somnambulists are not acting out some deep reflex unconnected to the mind. Nonetheless, "those faculties are operating at an unusually low level, one radically different from those of a person seized with deliberative capacities to act."[44] This means that the defendant is "incapable (on these occasions) of consciously directing her behaviour".[45] I want to suggest that all of this applies also to *Keal* and to other CLOC cases. The only difference is the precise way in which the faculties are "operating at an unusually low level": in particular, their operation is being *observed* by a conscious rather than unconscious mind. We should regard the morally relevant feature of these cases that causes them to fail the criminal law's voluntariness requirement as being their *subjective feeling of disconnection from conscious action*. (Un)consciousness *per se* is irrelevant, it is the lack of *subjective connection* to those higher faculties that seems to matter.

To support this claim, we may draw upon recent work by Stephen Garvey, who puts great emphasis on justifying insanity pleas by distinguishing between those who experience a "sense of agency", and those who do not.[46] Indeed, Garvey goes so far as to argue that this may be "the" key feature of all valid claims of insanity. Even if that more radical proposal is not endorsed, it does seem apt as an explanation of CLOC cases.

Garvey's view is that the best explanation of insanity's exemptive power is that insane defendants do not have a sufficient "sense of agency" over their behaviour. In

---

44 Simester, *Fundamentals, supra* note 1 at 227.

45 *Ibid*.

46 Stephen P Garvey, "Agency and Insanity" (2018) 66(1) Buff L Rev 123 [Garvey, "Agency and Insanity"]. Sinnott-Armstrong has offered an account along similar lines: Walter Sinnott-Armstrong, "A Case Study in Neuroscience and Responsibility" 52 NOMOS 194.

other words, it is the subjective experience of control that matters to responsibility, and the subjective alienation from behaviour that precludes blame:[47]

> The insane actor's mind and body commit the crime, but the mind and body committing the crime are not under his command. They've been commandeered. The choices his mind makes, the reasons moving his mind to make those choices, and the bodily movements resulting from those choices, are no longer experienced as his choices or his reason or his movements. An alien self is the one pulling the strings. It would therefore make no more sense to blame him for the crime resulting from those choices, reasons and movements than it would be to blame you or me. Blame presupposes a sense of agency, and insanity precludes blame because insanity defeats the sense of agency.

We do not need to adopt Garvey's view of the insanity offence wholesale to see the relevance of this theory to CLOC cases. It seems at least plausible that the best way to explain demonic possession cases[48] is in terms of that subjective sense of alienation. Doing so has the virtue of sidestepping many of the well-known problems associated with capturing the physiology of "lacking control". There is no need to define a threshold at which resistance to some internal urge becomes "too difficult". If the sense is genuinely of total alienation from control, and if this is a true description, backed by evidence, of those who feel they are in the grip of demonic possession, then this may be enough to justify a defence. What's more, we ought to categorise these cases as not merely ones of irresponsibility, but as instances of the profoundest kind of irresponsibility – that arising from a failure to satisfy the criminal law's voluntariness requirement. Indeed, as Garvey later states: "Blame and censure presuppose responsibility, responsibility presupposes agency, and agency presupposes a sense of agency."[49] Without that sense of agency, we are right to question whether a defendant can be said to have acted voluntarily.

C. *Simester's Too-Demanding Account of Involuntariness*

So far, I have proposed that the kinds of cases addressed by s 84(1)(*c*) can be understood as including involuntary behaviour notwithstanding that the defendant exhibits a form of conscious physical *agency* in his behaviour. We might say, in a sense, that he "chooses" to act as he does, inasmuch as there is a direct connection between his preferences and his behaviour: but that the choice itself is involuntary. Unfortunately, Simester's account of the criminal law voluntariness requirement seems inconsistent with this conclusion. For Simester, exercises of agency satisfy

---

[47] *Ibid* at 190.

[48] And likely many if not all CLOC cases. For example, it seems possible that cases of "overwhelming difficulty" (referred to earlier at in section III(a), where that sense of overwhelming difficulty fully alienates the defendant from his sense of agency, may also be (better) explained by that feeling of alienation. See also Simester's discussion of Winston in George Orwell's 1984: Simester, *Fundamentals, supra* note 1 at 423.

[49] Garvey, "Agency and Insanity", *supra* note 46 at 159.

the criminal law's voluntariness requirement, and even somnambulists come very close to satisfying that requirement, *save **only** for the fact that they are unconscious*. The reason he thinks that they are "almost" exercising their agency is that they are exercising their "higher faculties", the vaguely defined list of faculties that differentiate behaviour guided by the mind as opposed to mere bodily happenings. And he believes that it is the use of (or omission to use) these faculties that ultimately make conduct an exercise of agency. This view is not particularly controversial – there is considerable clinical and philosophical agreement that somnambulists' brains show many of the signs of waking, intentional behaviour. To avoid the bizarre implication that this means the law ought to treat sleepwalkers as having "acted" so as to ground basic responsibility, Simester invokes their unconsciousness. He writes: "this leads us to a further condition [of basic agency]: a person cannot be an agent in respect of a piece of behaviour unless she is conscious."[50] Therefore, "low-level" exercises of the "higher faculties" aren't good enough, and we assimilate "low level" exercises of them to "unconscious" exercises of them. I agree that somnambulists do not have even basic agency in relation to their unconscious "acts". However, if it is the unconsciousness of those acts that explains this, then it seems to leave the CLOC defendant on the hook for voluntary behaviour (though not necessarily for criminal responsibility, as involuntariness is not necessary for insanity). In a reformed criminal law without the insanity defence, defendants like Keal would not be entitled to invoke automatism as a failure of proof defence. This can be avoided if we move to a sense-of-agency analysis of Keal's involuntariness. That analysis would supply the deep explanation of Keal's defence, whether that defence was legally provided, as currently, under insanity doctrine or, in a world where the defence was abolished, via automatism.

It may be wondered why, on Simester's view, the somnambulist's mere incapacity to do otherwise than act in the way that he did is not enough to render him involuntary. After all, even if we grant that there was an exercise of the higher faculties in such cases, there isn't much the somnambulist can do about it. This sensible move is in fact available to those who explain voluntariness in terms of "incapacity to do otherwise" as opposed to by reference to exercises of agency. But there is reason to believe that the same move is not available to Simester. This is because Simester combines both approaches to voluntariness – voluntariness as capacity to do otherwise and voluntariness as exercise of agency – to achieve a more subtle, pluralist view. In many ways, this is an attractive approach, which plausibly explains a range of test cases.[51] As Simester rightly shows, sometimes a defendant will meet the voluntariness requirement even though he did not exercise his agency. Simester gives the example of a person who sneezes. In this case, the person's *capacity* to stop that behaviour (insofar as they have it – not all sneezes can be stopped!) is what grounds their voluntariness. Yet we would not say that they were the "author" of the sneeze. Similarly, where D exercises his agency to move his body, but in circumstances

---

[50] Simester, *Fundamentals, supra* note 1 at 227.

[51] Simester sums up his "disjunctive" account of voluntariness in the following test: "D's behaviour is voluntary (L) — that is, it satisfies the law's voluntariness requirement — if either that behaviour is non- involuntary (in the sense that D has the deliberative capacity not to behave as she does) or that behaviour is an instance of her agency.": Simester, *Fundamentals, supra* note 1 at 193.

where he had no capacity to do otherwise – Simester gives the example of a stronger person standing by with the intention to grab and move D's body in the same way – this *also* counts as criminal law voluntariness for Simester. D has at least basic responsibility in relation to that movement. Again, this approach seems exactly right where the incapacity to do otherwise than act in a specific way is caused by the threat of an external, physical force. This was Simester's focus in the example.

Yet, when applied to the case of the somnambulist, we can see that the otherwise sensible pluralist approach seems to indicate that the somnambulist's behaviour is voluntary, because mere incapacity to do otherwise is not enough to deny voluntariness when the "higher faculties" are engaged. This is why Simester reaches for the conscious/unconscious line to sort out those exercises of higher faculties that satisfy the voluntariness requirement, and those that do not. And that move is what condemns Keal, not himself a somnambulist, to the voluntary side of the divide. If the sense of agency account can be made to work, that implication no longer follows. There is no obvious reason why Simester's account of involuntariness could not be modified to allow for it.

This section has not offered a full argument for the sense-of-agency analysis of cases like Keal's. As with the other theories discussed, answers to difficult empirical questions of physiology will need to be provided before we can determine whether the argument is sustainable. Equally, we have reason for caution before introducing a test that puts so much emphasis on the subjective feelings of the defendant. Will it be practically justiciable: what counts as a sufficiently "authentic" feeling of lost agency, and how will it be proven? The potential for abuse is clear. One strength of Simester's invocation of unconsciousness was that it avoided some of that vagueness. Nonetheless, the alternative approach canvassed here strikes me as plausible and one that avoids certain pitfalls. I have argued that it has the potential to play an important and practically significant role in understanding the insanity defence.

VII. Conclusion

Simester's taxonomy of defences is informed by the subtle interrelation of foundational principles like moral responsibility and culpability. Careful attention to those principles shows why it would be a mistake to categorise irresponsibility defences like insanity as excuses. Yet irresponsibility defences themselves should be separated into voluntary and involuntary varieties. Doing that is no easy task. I have suggested that cases like Keal's are best thought of as involving involuntary behaviour. That analysis may well be controversial. But, whether or not it is accepted, my point is also that its acceptance has crucial implications for determining whether the insanity defence is "inculpatory" or "exculpatory" for such defendants; and for determining which doctrines should be available to them if further statutory reform to the unsoundness of mind defence is pursued.