

CAPACITY, MORAL RESPONSIBILITY AND THE CRIMINAL LAW

*James Slater**

INTRODUCTION

This article is concerned with the conflict between two theories of moral responsibility for wrongdoing, one of which I shall term the shallow capacity theory and the other the self-control theory. This conflict is of interest for two reasons. First, and fundamentally, it is important from the perspective of moral philosophy: in this regard, I will argue that the shallow capacity theory is incomplete, and as a result inferior to the self-control theory, which offers a complete account of moral responsibility for wrongdoing. Secondly, given the criminal law's interest in moral responsibility and blame, I will argue that the self-control theory offers two important insights for the criminal law, insights that the shallow capacity theory does not provide. First, it offers the most accurate understanding of the moral significance of killing under provocation, and thus the best framework for understanding the partial defence of provocation. Secondly, it demonstrates that there is a need in the criminal law for a defence based on radical impairment of an agent's capacity for self-control, and in so doing offers a vital insight into the notion of a partial denial of moral responsibility. It should be noted that these insights for the criminal law emerge from those features of the self-control theory that make it superior to the shallow capacity theory as a theory of moral responsibility.

This article will proceed as follows. In the first section, I will explain what I mean by moral responsibility for wrongdoing, and in so doing outline the important distinction between exemptions and excuses. In the second section, I will outline and contrast the shallow capacity and self-control theories, and in so doing describe the conflict between them. In the third and fourth sections, I will show how the shallow capacity theory suffers from two fundamental flaws, flaws not shared by the self-control theory. I will also explain the insights offered by the self-control theory for the criminal law in these sections. In the fourth section, I will conclude by offering some general comments on the conflict between the shallow capacity theory and the self-control theory and how the debate can be taken further.

* BA (Sussex), LLM (Harvard), PhD (Birmingham), Lecturer in Law, the University of Buckingham. I am indebted to Stephen Shute, Andrew Simester, John Gardner and Susan Edwards for their insightful comments on earlier drafts.

MORAL RESPONSIBILITY, EXEMPTION AND EXCUSE

When I talk of moral responsibility for wrongdoing, I mean the necessary and sufficient conditions for entering the process of moral evaluation that may issue in the ascription of moral blame for wrongdoing. Moral responsibility is therefore preliminary to the process of moral evaluation. Whether blame is eventually allocated depends on that process of moral evaluation, so it is important to bear in mind that morally responsible wrongdoers are not necessarily blameworthy. Thus moral responsibility explains how, why, and to what extent a wrongdoer is a legitimate object of the moral evaluation that precedes, and may issue in, the ascription of blame.

The distinction between moral responsibility and the evaluation upon which the final allocation of blame depends means that exculpation for wrongdoing takes two forms.¹ First, a wrongdoer may deny her moral responsibility. Such an agent denies that she is, fully or partially, a legitimate object of the moral evaluation that precedes the allocation of blame. This exculpation will take the form of a full or partial exemption. As we will see below, examples of exemptions include insanity and infancy. The idea is that the insane and infants are *exempt* from the moral evaluation that can lead to the allocation of blame because they lack certain requirements necessary for that moral evaluation to make sense. Secondly, a morally responsible wrongdoer may be able to exculpate herself by showing that she is not blameworthy within the terms of the moral evaluation. This will take the form of a full or partial excuse. I will argue below that anger-inducing provocation is best conceptualised as an excuse, because the exercise of the provoked wrongdoer's capacity for self-control, the possession of which makes her morally responsible, is morally evaluated, and blame allocated when it is exercised poorly.

To summarise, exemptions deny, fully or partially, the appropriateness of engaging in the moral evaluation that may issue in blame, whereas excuses accept that the wrongdoer is (more or less) a legitimate candidate for that

¹ I am assuming that the wrongdoer is not justified. Exculpation in the form of justification is the province of normative theory, whereas this paper is concerned with ascriptive theory. As Joseph Raz explains, “[n]ormative theory is primarily concerned with establishing what people ought to do”: *Practical Reason and Norms* (Oxford: Oxford University Press, 2nd edn, 1999) p 11. Ascriptive theory, in contrast, is not concerned with what agents ought to do, but with the allocation of praise or blame to agents for doing, or not doing. As such, as John Gardner points out, it does “not govern what should be done by anyone, but whether and how we should count what people have done when we are judging them”: “Criminal Law and the Uses of Theory: A Reply to Laing” (1994) 14 *Oxford Journal of Legal Studies* 217 at 220.

moral evaluation, but show that the ascription of blame is not appropriate, or not fully appropriate, as its conclusion.

THE SHALLOW CAPACITY THEORY AND THE SELF-CONTROL THEORY

The Conflict between the Shallow Capacity Theory and the Self-Control Theory: A Preliminary Outline

According to the shallow capacity theory, an agent will be morally responsible for her wrongdoing, and therefore a legitimate candidate for moral evaluation and the potential allocation of blame for that wrongdoing, if two requirements are satisfied. First, the agent must be the author of the wrongdoing. This means that the wrongdoing must spring from the agent in such a way that it belongs to her. Secondly, the agent's wrongdoing must be intelligible. This means it must emerge from the agent's intelligible understanding of the pertinent factual and normative features of her situation. In contrast, the self-control theory holds that there is more to moral responsibility than authorship and intelligibility, though it accepts, with an important modification where authorship is concerned, that these are necessary features of moral responsibility. According to the self-control theory, the agent must not only have a capacity to act intelligibly: she must also possess a capacity to act in accordance with the reasons she deems best in light of her conception of the good. This capacity, which I call the capacity for self-control, includes a volitional element, by which I mean an active power for self-determination where action is concerned.² Possession of this capacity, according to the self-control theory, is a necessary condition of an agent's moral responsibility.

The shallow capacity theory explicitly rejects the need for the capacity for self-control. As far as the shallow capacity theory is concerned, an agent is morally responsible so long as she is a springboard for rational action: a volitional capacity to control action in light of better judgment is unnecessary. Over the next three sub-sections, I will tease out the exact nature of this conflict as follows: first, I will examine in greater detail the shallow capacity's requirements for authorship and intelligibility, and articulate the exemptions and excuses that flow from these requirements; second, I will describe in greater detail the capacity for self-control as required by the self-

² I am indebted to the work of R Jay Wallace for this notion of a volitional power over action: see, for example, "Addiction as Defect of the Will: Some Philosophical Reflections" (1999) 18 *Law and Philosophy* 621-624 and "Three Conceptions of Rational Agency" (1999) 2 *Ethical Theory and Moral Practice* 217-242.

control theory; third, I will describe where the two theories converge, and where, crucially, they part.

The Shallow Capacity Theory: A Closer Look

As stated above, the shallow capacity has two, and only two, requirements in order for an agent to be morally responsible for her wrongdoing: first, the requirement that the agent be the author of the wrongdoing; second, the requirement that the wrongdoing be intelligible, that is to say based on a rational interpretation, and response to, the agent's situation.

The shallow capacity theory's approach to authorship for wrongdoing is characterised by two features. First, the notion that the wrongdoing must be linked to the agent in such a way that it can be said that the wrongdoing springs from the agent in the required sense.³ Thus one way an agent can be exempt from moral responsibility according to the shallow capacity theory is if this link between the agent and the wrongdoing is broken. This requirement explains the exculpatory effect of various phenomena grouped under the heading of automatism in Anglo-American criminal law, such as bodily movements during a seizure, as these break the link between the agent and her wrongdoing. In such cases, the agent is not the author of the action: there is no connection between her values, beliefs, desires etc and her physical movements.

The second feature of the shallow capacity's approach to authorship is its refusal to compartmentalise the agent at the psychological level, with the result that all cognitive and emotional features of the agent constitute the agent, however transient they may be, or hidden from consciousness.⁴ The result is that wrongdoing that is in the required sense linked to those features is wrongdoing for which the agent is morally responsible.⁵ In the same vein, the agent cannot identify herself exclusively with her better judgment, and thereby deny moral responsibility for wrongdoing that is the product of weakness of will,⁶ or identify herself exclusively with her character, and thereby deny moral responsibility for wrongdoing that is out of character.⁷

³ Eugene Schlossberger *Moral Responsibility and Persons* (Philadelphia: Temple University Press, 1992) p 45. See also Robert M Adams "Involuntary Sins" (1985) 94 *Philosophical Review* 3 at 25.

⁴ Schlossberger, *ibid* pp 34 and 44 and generally Adams *ibid*.

⁵ Schlossberger, *ibid* p 41, p 44 and p 67, Adams, *ibid*, and Andrew Simester "Can Negligence be Culpable" in Jeremy Horder (ed) *Oxford Essays in Jurisprudence 4th Series* (Oxford: Oxford University Press, 2000) p 85, p 98.

⁶ Schlossberger, *ibid* p 41 and p 67.

⁷ Schlossberger, *ibid* p 44.

In addition to authorship, the shallow capacity theory also requires that the agent possess a fundamental capacity for rational action at the time of the wrongdoing, with the result that the wrongdoing is itself intelligible, even if morally deficient. Thus, in addition to denials of authorship, an agent can deny her moral responsibility according to the shallow capacity theory by denying this capacity.⁸ The result is that the insane and infants are exempt from moral responsibility as they lack this core rational capacity. The shallow capacity theory also accepts that possession of this rational capacity is a matter of degree, and thus children of a certain age, whose rational capacity is partially developed, are, as a result, partially morally responsible.⁹

The reason why the shallow capacity theory can be characterised as shallow is because, once the agent is the author of her wrongdoing, there is only one further condition for moral responsibility: that the wrongdoing is intelligible. Thus wrongdoing that can be explained in terms of an intelligible evaluation of, and reaction to, moral reasons, means the wrongdoer is a legitimate candidate for blame, without more. In turn, there are only two ways to deny moral responsibility with the shallow capacity theory: either to break the link between the agent and the wrongdoing, and thereby deny authorship, or deny the agent was capable of rational action *tout court*.

Further understanding of the shallow capacity theory is obtained through an examination of its approach to excuses. It is to be remembered that once a wrongdoer is morally responsible, she is a legitimate candidate for moral evaluation in light of her wrongdoing, with the allocation of blame dependent on the conclusion reached as a result of that evaluation. According to the shallow capacity theory, moral responsibility depends exclusively on the capacity to act intelligibly. The result of this approach is that the only subject matter upon which blame and excuses can be allocated is the moral quality of the agent's intelligible motives for her wrongdoing.¹⁰ Thus, if the wrongdoing expresses morally defective values, then the agent is (more or less) blameworthy. If the wrongdoing expresses (some) morally appropriate values, then the agent may not be fully blameworthy.¹¹ In such cases, a full or partial

⁸ Gardner "The Gist of Excuses" (1998) 1 *Buffalo Criminal Law Review* 575 at 588-589 and "Justifications and Reasons" in Andrew P Simester and Anthony T H Smith (eds) *Harm and Culpability* (Oxford: Clarendon Press, 1996) pp 103-129, at pp 121-122. Also Schlossberger, *ibid* p 144.

⁹ Schlossberger *ibid* 74-76.

¹⁰ The shallow theory also purports to encompass moral responsibility for negligence, a form of wrongdoing that may appear unmotivated, by interpreting the negligence as constituting or expressing a morally defective attitude, see Schlossberger *ibid* p 109 and Simester above n 5 generally.

¹¹ Nicola Lacey summarises this approach to excusatory exculpation as follows: "The focus, crucially, is upon the quality of the attitude manifested *in the defendant's*

excuse may be appropriate.¹² This can be illustrated with chosen wrongdoing and duress.

Let us take two hypothetical soldiers, Michael and Paul, captured by the enemy. Michael is threatened with torture and execution unless he reveals certain important military secrets, and proceeds to do so. Meanwhile, Paul is threatened with being tickled unless he reveals the same secrets, which he also proceeds to do. Michael's judgment that terrible pain is worth avoiding and that his life is worth saving has certain positive moral features: a subtle moral analysis of his motivation reveals certain morally acceptable dimensions, and is not indicative a person whose priorities are ranked entirely incorrectly. He can be partially excused as a result. By contrast, Paul's choice shows that he values not being tickled, a trivial event, over the fundamental interests of others, and this is an incorrect moral judgment. Paul is therefore fully blameworthy.¹³

conduct, evaluated in the light of (a generous interpretation of) the context in which it occurred [her emphasis]", see her "Partial Defences to Homicide" in M Mitchell and A Ashworth (eds) *Rethinking English Homicide Law* (Oxford: Oxford University Press, 2000) p 107, p 119. Lacey calls this the "reasons view" of criminal defences, as it focuses on the agent's reasons for violating the criminal law, with exculpation allocated when those reasons exhibit certain morally acceptable features. She also points out how those who subscribe to this view distinguish such excusatory defences from exemptions such as insanity and diminished responsibility (p 117), as those benefiting from the latter defences "[are] not capable of acting responsibly in the given sense" (p 119). Lacey is describing the position of those who subscribe to the shallow capacity theory.

¹² For an argument that there is no such thing as an excuse that fully extinguishes blame for wrongdoing, see N Dahl "'Ought' and Blameworthiness" (1967) 64 *Journal of Philosophy* 418.

¹³ I have adapted this example from Schlossberger above n 3 p 106. It is important to note that Gardner conceives of excuses as instances of "indirect rational explanation", whereby the agent does not directly justify her wrongdoing, but rather the beliefs, emotions or desires that led to it. According to Gardner, duress and provocation therefore excuse when the wrongdoer can point to good reasons for her fear or anger. It is a purely evaluative approach that focuses on the moral quality of the agent's rational reasons for feeling as she does: see above n 8, especially p 119-120, and also "The Mark of Responsibility" (2003) 23(2) *Oxford Journal of Legal Studies* 157 and "No Provocation Without Responsibility: A Reply to Mackay and Mitchell" (2004) *Criminal Law Review* 213, especially at 214-215. Gardner's theory is a version of what Lacey calls the "reasons view" of criminal law defences (see above n 11).

The Capacity for Self-Control as Conceived by the Self-Control Theory

The capacity for self-control as conceived by the self-control theory requires, as a necessary background feature, the possession by the agent of an intelligible conception of the good. This acts as a general blueprint by which the agent assesses and plans her conduct. Such a conception is made up of the agent's diachronically stable, coherent and more or less mutually consistent higher order moral and prudential values. Key to such values is their importance to the agent: as N J H Dent explains, a rational agent "will come to some sense of which [various concerns] have a more enduring significance to him, a sense of their significance which will survive the occasional disappearance or overlaying of any passionately engendered sense of their importance. One could say that he will come to an appreciation of their value."¹⁴ Conceptions of the good therefore consist of those values, aims and perspectives that, as Gary Watson puts it, "in a cool and non-self deceptive moment [the agent] articulates as definitive of the good, fulfilling, and defensible life."¹⁵

What I mean by the capacity for self-control is the agent's capacity to govern her conduct in accordance with her conception of the good whatever the desires she is experiencing.¹⁶ The self-control theory therefore rejects the notion of the agent as the passive object of the various desires within her psychology, blown one way or another at any given time by the most causally powerful amongst them, but rather conceives of the agent as possessing a distinct and active capacity to control her actions in light of her deliberative judgements about what she has most reason to do. It is this capacity to control conduct whatever the desires the agent is experiencing that gives the capacity for self-control its active, or volitional, component. This approach to the genesis of action is therefore in contrast to what may be termed the hydraulic account of human motivation, whereby the agent is always at the mercy of her most powerful desires, and whether or not she exercises self-control purely the fortuitous result of a match between her most powerful desire and her better judgment.¹⁷

¹⁴ *The Moral Psychology of the Virtues* (Cambridge: Cambridge University Press, 1984) p 111.

¹⁵ Gary Watson "Free Agency" in G Watson (ed) *Free Will* (Oxford: Oxford University Press, 1982) pp 96-110, p 105. It is worth noting that a rational agent will also accept that her values and goals may develop or change in light of new experiences, information and argument.

¹⁶ I am using the term 'desire' widely, to include emotions, urges and impulses.

¹⁷ Once more I draw the reader's attention to the work of Wallace, above n 2, in particular 630-638 of "Addiction as Defect of the Will: Some Philosophical Reflections", where Wallace contrasts what he calls the volitionalist and hydraulic

The capacity for self-control is called upon whenever the agent experiences desires that prompt conduct inconsistent with her conception of the good. Exercising self-control therefore means the agent exploiting her capacity for self-control in order to act, or refrain from acting, as required by her conception of the good in the face of incontinent desires. In turn, when an agent fails to exercise self-control, for example under the influence of addiction, anger or merely temptation, she acts (or fails to act) for reasons that she rejects in light of her conception of the good, or, if not rejects, believes to be outweighed by other reasons.¹⁸

The capacity for self-control is therefore more than a mere capacity for intelligible action, since it actively empowers the agent to put her settled higher order values into practical effect whatever the emotions, desires etc she is experiencing. It is worth noting that the capacity is complex, and involves more than a purely volitional component to act on better judgment once practical resolutions have been reached. It also involves an active capacity to prevent emotion and the like from corrupting practical reason, thereby ensuring that the agent's practical resolutions and intentions are in accordance with better judgment. Further understanding of this feature of the capacity will emerge below, when I discuss the exercise of self-control in the context of provocation and the emotion of anger.

Agreement and Conflict between the Two Theories

There is a measure of agreement between the shallow capacity and self-control theories. This is because the self-control theory does not reject the exemptions that flow from the shallow capacity's requirements for authorship and intelligibility. Thus the self-control theory accepts that certain exemptions, such as reflex movements or bodily movements during seizures, exempt by breaking the link between the agent and her wrongdoing. After all, it makes no sense to evaluate an agent morally for wrongdoing with which she is in no way connected. The self-control theory also accepts that agents who are fundamentally incapable of rational action should be exempt. Since moral responsibility involves a moral evaluation of the agent in light of her wrongdoing, if the wrongdoing is not an expression of a rational (though perhaps morally defective) set of values, moral evaluation, once more, makes

conceptions of human motivation, and rejects the latter. For another argument that treats self-control as a distinct faculty that is exercised by the agent, see Richard Holton "How is Strength of Will Possible?" in S Stroud and C Tappolet (eds) *Weakness of Will and Practical Irrationality* (Oxford: Clarendon Press, 2003) pp 39-67.

¹⁸ In the context of negligence (inadvertent wrongdoing), this means the agent does not wish to bring about, or risk, the harm she inadvertently causes or risks.

no sense.¹⁹ Where excuses are concerned, there is also some agreement, since the self-control theory accepts that some excuses may depend exclusively on an analysis of the moral quality of the agent's motivation.

The conflict between the two theories is due to the fact that the self-control theory considers the shallow capacity theory incomplete; according to the self-control theory, moral responsibility also requires, in addition to the capacity for intelligible action, possession of the capacity for self-control as outlined in the previous section. This requirement is something that the shallow capacity theory explicitly rejects.²⁰ This disagreement is closely linked the shallow capacity theory's approach to authorship. As stated above, the shallow capacity theory adopts a global and wholly inclusive view of the agent, embracing all her emotional and cognitive features. It also holds that once wrongdoing is linked to those features in the required sense, the agent is morally responsible for that wrongdoing, without more. Even if the agent is incapable of controlling her conduct in light of her better judgement, this is irrelevant to her moral responsibility, because the shallow capacity considers such action an expression of the agent's agency. It is for this reason that

¹⁹ Schlossberger above n 3 p 144 ("The insane are not responsible, I suggest, because we cannot tell a coherent story about them as moral evaluators that makes sense of their lives. We cannot attribute to them stable attitudes, coherent goals, values to which they are committed, and so forth. We cannot regard their acts as expressing a moral framework, whether good or ill"). See also Gardner above n 8 p 121 and above n 8 at 589 ("But the focus on making sense of people's actions in the light of their reasons rightly brings to the surface the important point that those whose reasoning *can't* be made sense of in this way, whether because of profound mental illness or infancy or sleepwalking or (on some interpretations of it) post-hypnotic suggestion, are not responsible for their actions and therefore need no excuses for what they do").

²⁰ Schlossberger *ibid* p 41 and p 67; Adams above n 3 pp 12-13, Simester above n 5 p 98 and Gardner *ibid* 581-582, 584-585 and 589. Gardner is skeptical about the whole notion of a capacity for self-control, because he believes there is no such thing as unexploited capacity, or, to put it colloquially, capacity to spare. All behaviour is at the limit of whatever that agent is capable of at the moment action. This is not to say that Gardner rejects the need for a second-order rational capacity before one is morally responsible. He believes, as explained in an e-mail to me, that an agent's actions need to be sensitive to her reasons for acting, and her reasons for acting in turn need to be sensitive to the facts (i e to her reasons for holding them to be reasons). In other words, Gardner believes that the agent needs to be capable of evaluating and selecting the reasons for which she acts (for that is how they become her reasons). Where Gardner and I disagree is that Gardner does not believe that the agent needs any kind of active or volitional control over the reasons for which she acts. By contrast, I believe that agents do have capacity to spare, capacity that they can fail to exploit, and that possession of this capacity in the context of self-control is a necessary feature of moral responsibility for wrongdoing.

weakness of the will poses no problem for the shallow capacity theory, because the agent is identified with the values expressed in her wrongdoing, even though they are rejected in light of her better judgment.²¹ The self-control theory, in contrast, argues that self-control does raise questions for moral responsibility, and that a complete understanding of moral responsibility, and hence exemptions and excuses, requires a theory that treats possession of the capacity for self-control as a necessary condition of moral responsibility. It is for this reason that it considers the shallow capacity theory incomplete.

I will expose the incompleteness of the shallow capacity theory over the next two sections of this article by discussing two flaws that flow from its rejection of the need for the capacity for self-control. The first of these flaws manifests itself in the context of the partial excuse of provocation, the other in the context of exemptions.

THE SHALLOW CAPACITY THEORY'S FIRST FLAW: AN INCOMPLETE ACCOUNT OF PROVOCATION AS EXCUSE

This section of the article will proceed in two stages. First, I will argue that the shallow capacity theory is flawed because it cannot give a complete account of why anger-inducing provocation should sometimes partially excuse murder, in contrast to the self-control theory, which can.²² The reason for this flaw lies in the shallow capacity theory's refusal to see that the exercise of self-control lies at the heart of the correct conceptualisation of the partial excuse. Dan M Kahan and Martha C Nussbaum's approach to the partial defence of provocation, a purely evaluative theory that rejects the relevance of self-control and therefore reflects the approach of the shallow capacity theory, will be used to expose this flaw.²³ Second, I will assess the criminal law in light of my argument that the exercise of self-control lies at the heart of the partial defence of provocation, including a comment on the Law Commission's recent proposals for reform of the defence.²⁴

²¹ See Schlossberger *ibid* p 67.

²² References to provocation from now on should be taken to refer to provocation that triggers anger.

²³ "Two Conceptions of Emotion in Criminal Law" (1996) 96/2 *Columbia Law Review* 269.

²⁴ Law Commission *Murder, Manslaughter and Infanticide* Law Com No 304 (London: TSO, 2006) ('Law Com 2006'), Part 5.

Exposing the Flaw in the Shallow Capacity Theory's Conception of Provocation as Excuse

In addressing provocation and the emotion of anger, Kahan and Nussbaum treat cognition, in the form of an agent's evaluation that she has been wronged and her evaluation of what she should do in response to that wrong, as the predominant component of the emotion of anger.²⁵ In turn, they argue that provocation's excusatory nature depends *exclusively* on an assessment of the agent's evaluation that she has been wronged. If that evaluation involves the agent perceiving violations of certain fundamental moral values, for example the rape or killing of a loved one, then, according to Kahan and Nussbaum, the agent can be justifiably enraged and partially excused for her intentional killing,²⁶ and thereby be convicted of voluntary manslaughter.²⁷ However, they recognise, correctly, that there is something defective about the evaluation embodied in the agent's anger, in that she has failed to give the correct moral weight to the value of the provoker's life and the importance of lawful resolution of disputes. They conclude that it is this defect that results in provocation acting as a *partial* excuse.²⁸

Key to Kahan and Nussbaum's argument is that anger's impact on the exercise of self-control is irrelevant to why provocation can excuse; in this regard, they state, "The moral status of emotions [] is determined by the quality of the values they express, not by their effect on volition."²⁹ Though I agree that provocation is correctly conceived as a partial excuse, it seems to me that this purely evaluative approach, which treats questions of self-control as irrelevant, is incomplete. In turn, when the incomplete nature of the purely

²⁵ This reflects their "evaluative conception" of emotion generally: above n 23, at 285-297.

²⁶ Strictly speaking, the defendant might only have intended grievous bodily harm under the English definition of murder.

²⁷ Above n 23, at 305-323, in particular at 313. For a similar argument, see V Nourse "Passion's Progress: Modern Law Reform and the Provocation Defense" (1997) 106 *Yale Law Journal* 1331.

²⁸ Kahan and Nussbaum, *ibid* 313. It is worth noting Gardner's argument that all excuses, full or partial, suffer from some moral defect, and that it is this defect that distinguishes not full excuses from partial excuses, but excuses of any kind from justifications; see, for example, "Justifications and Reasons" above n 8 pp 103-129.

²⁹ *Ibid* at 343. Though the following quote refers to duress and the emotion of fear, it captures the essence of Kahan and Nussbaum's approach to anger and their general thesis that control issues are not relevant: "...duress exculpates not because (and when) threats vitiate a person's moral agency, but because (and when) a person's fear expresses a rational and morally appropriate assessment of her circumstances" (336).

evaluative approach is explained, it exposes a flaw in a theory of moral responsibility that treats self-control as irrelevant to moral responsibility.

Kahan and Nussbaum's theory is incomplete because the moral defect in the provoked agent's anger is fatal to excuse based exclusively on the moral quality of that anger. The reason emerges from the backward-looking quality of the provoked agent's reasons for acting. Unlike cases of self-defence or duress, where the threatened agent acts prospectively to prevent harm or death to a loved one or herself, provoked agents are, as Jeremy Horder explains: "inflicting injury in response to a perceived wrongdoing that is now a part of history, albeit recent history."³⁰ In other words, the provoked agent is not acting to protect her autonomy or the autonomy of another, but to impose private punishment.³¹ The problem is that killing in the name of private punishment is, without more, fully culpable. The reason why backward and forward-looking reasons for acting are distinguishable in this context is derived from the criminal law's attitude to self-help combined with the moral quality of the agent's reasons for acting as she does.

With criminal law justifications/excuses based on forward-looking reasons, such as duress and self-defence, the agent is presented with a more or less pressing need to take action in the name of self-preservation or the preservation of another. This need not manifest itself in an immediate temporal need for action, but nevertheless means that the agent must act there and then; an example would be a lone and isolated hostage told she will be executed in 24 hours and presented with an opportunity to attack her kidnapper with 12 hours to go. The agent's actions in such cases have a special moral purpose: to restore or maintain her own autonomy, or the autonomy of another, in the face of a threat that is, in some sense, illegitimate.³² This moral purpose is one shared by the State, with the result

³⁰ "Autonomy, Provocation and Duress" [1992] *Criminal Law Review* 706 at 709. For a similar argument, see A von Hirsch and N Jareborg "Provocation and Culpability", in F Schoeman (ed) *Responsibility, Character and the Emotions: New Essays in Moral Psychology* (Cambridge: Cambridge University Press, 1987) pp 241-255, pp 245-248; in this regard, von Hirsch and Jareborg note: "The claim of extenuation comes not from currently threatened interests, but from the defendant's sense of grievance" (p 248).

³¹ Horder, *ibid* 709-711.

³² The above discussion distinguishes between provocation, on the one hand, and self-defence and duress, on the other, on the basis that action in the former is based on backward-looking reasons and action in the latter is based on forward looking reasons. However, it is accepted that self-defence arguably has a backward looking element, at least in ideal form, in that there must be an aggressor who is legitimately harmed so that her attack can be thwarted. However, acknowledging this does not detract from the argument in the text above, because that argument does not hold that

that, if the State had an opportunity to neutralise the threat, it would take that opportunity. Furthermore, and crucially, the State's response would be justified because the threat is illegitimate, and its response can be tailored to remove the threat efficiently due to its overwhelming power. However, if the State cannot intervene, the private citizen, based both on need and also the illegitimacy of the threat faced, is justified in, or at least excused for, neutralising the threat where self-defence is concerned,³³ or avoiding the threat where duress is concerned.³⁴ Conversely, if there is an opportunity for

acting for backward looking reasons by definition cannot sustain an excuse, rather than the backward looking reasons present in cases of provocation cannot do so if provocation's excusatory potential depends purely on the moral quality of the agent's anger. In provocation, those backward looking reasons are purely punitive, whereas in self-defence, they define when acting for forward-looking reasons is justified. For a sophisticated discussion of when killing in self-defence is justified, see Suzanne Uniacke *Permissible Killing; The Self-Defence Justification of Homicide* (Cambridge: Cambridge University Press, 1994).

³³ The force used to neutralize the threat in self-defence must be reasonable in light of the threat or the honestly perceived threat: see *Scarlett* [1994] 98 Crim App R 290 and *Owino* [1996] 2 Cr App R 128. There is much debate as to whether an agent is justified or merely excused when she mistakenly but reasonably believes she is under attack, and then reacts proportionately to the perceived attack. Some argue that the agent is justified in such circumstances; see, for example, Joshua Dressler "New Thoughts About the Concept of Justification in the Criminal Law" (1984) 32 *UCLA Law Review* 61 at 92-95 and Kent Greenawalt "The Perplexing Borders of Justification and Excuse" (1984) 84 *Columbia Law Review* 1897, at 1907-1911. Others argue that the agent is merely excused; see, for example, George Fletcher *Rethinking Criminal Law* (Little, Brown and Company, 1978) pp 762-769 and "The Right and the Reasonable" (1985) 98 *Harvard Law Review* 949 at 972-973. The fact that English law allows unreasonable but honestly believed mistakes as to an attack to lead to an acquittal (with the exception of those generated by voluntary intoxication: see *O'Grady* [1987] 3 WLR 321 and *O'Connor* [1991] *Criminal Law Review* 135) is the product of a certain doctrinal view flowing from the requirement for *mens rea* (see Gardner above n 8 p 121). However, if the mistaken belief in an attack is the product of gross negligence and death results, the defender is guilty of gross negligent manslaughter.

³⁴ There remains a question mark over whether the defence of duress is a justification or an excuse. This arguably depends on the nature of the crime specified in order to avoid realisation of the threat. However, given that the required threat is of death or serious personal injury (see *Graham* (1982) 74 Cr App R 235), it is submitted that many instances of duress will be justifications. In any event, the threat represents an illegitimate restriction on the autonomy of the person threatened, a quality duress shares with self-defence. It is worth noting that English law is not clear on whether the defence is lost when the threatened agent has an honest but unreasonable belief

the State to neutralise the threat, the agent should allow the State to do so; in such circumstances, private action, no longer necessary, ceases to be justifiable or excusable.³⁵ It is for this reason, in part, that the criminal law imposes the requirements of imminence for duress and necessity for self-defence. Thus, to summarise, when the agent cannot rely on the State to protect her autonomy or the autonomy of another, actions taken to secure that autonomy are justifiable in some cases and excusable in others, and this is reflected in the criminal law.

The above arguments do not apply with provocation. Where self-help is concerned, unlike duress and self-defence, there is no practical need for any action in order to protect autonomy, but only an impulse for a response based on grievance. Perhaps this would not matter if the agent's reasons for killing were morally highly compelling in light of the reasons against killing, but they do not measure up in light of those very powerful reasons. It is for this reason that the criminal law does not share in the values expressed in the act of killing, because the State does not condone punishment of such severity, whatever the provocative act or event.³⁶ In light of these arguments, private individuals are expected to refrain from killing purely in the name of punishment, and remain fully culpable if they do so, however grave the wrong committed against them. This is not to say that the presence of a morally grave provocation does not have an excusatory role to play: naturally, no excuse should be available if the provocation is not sufficiently serious.³⁷ But an additional excusatory factor requires articulation if we are to understand why we excuse in light of the provoked agent's failure to accord the correct moral weight to the value of the agent's life and the importance of the lawful resolution of disputes. That necessary additional excusatory factor is the difficulty that sufficiently intense anger creates for self-control, because that difficulty, as I will explain further below, allows us (partially) to overlook the agent inflicting such severe punishment at her own initiative. However, before

that a threat is being made: see *Martin* [2000] 2 Cr App R 42 and *Safi* [2003] EWCA Crim 1809.

³⁵ When the agent fails to exploit an opportunity for State intervention, more emphasis could be placed on fear by both the self-defence and duress defences in order to excuse the fact that the agent does not allow for that State intervention. In this sense, the effect of fear on self-control in the context of self-defence and duress should be acknowledged by the criminal law in the same way, as I will argue in the text above, that it acknowledges the effect of anger in the context of provocation.

³⁶ Perhaps under certain circumstances, the death penalty is morally justified for certain crimes, though this is highly debatable. Even so, the infliction of such punishment should be, in any event, the exclusive province of the State.

³⁷ Thus the gravely provoked agent's reasons for acting are merely a necessary condition for an excuse, but not sufficient of themselves.

I turn to the exact nature of this difficulty for self-control, it is necessary to outline how self-control functions as a capacity in this particular context.

Anger in the context of provocation is the experience of more or less intense physiological arousal, commonly known as affect, triggered by an evaluation that a wrong has been committed against the angry agent or someone close to her.³⁸ Anger creates problems for self-control because, as Stephen Gough explains, “[it] modifies perspective, directing focus towards (existing) reasons to confront and fight at the expense of reasons to bite one’s tongue, walk away or otherwise act sensibly.”³⁹ Anger is able to modify perspective in this way because its affective dimension makes the agent uncomfortable, sometimes acutely, in not contemplating retributive action.⁴⁰ The discomfort inclines the agent towards the choice to harm the provoker because when that choice is entertained and acted on, the anger, and hence the

³⁸ It is a commonly held thesis that the evaluation of such wrongdoing not only triggers anger, but also partly constitutes the emotion of anger: see, for example, Kahan and Nussbaum, *above* n 23 at 282-4; Patricia S Greenspan *Emotions and Reasons: An Enquiry Into Emotional Justification* (New York: Routledge, 1988) pp 48-55; and Timothy Macklem and Gardner “Provocation and Pluralism” (2001) 64 *Modern Law Review* 815 at 820. Deciding this question is not necessary for the purposes of this paper, though it is accepted that the degree of an agent’s anger, when that agent is emotionally healthy, is proportionate to her evaluations: see n 42 below. The cognition-affect debate with regard to emotion is well summarized in chapter 1 “The Cognition-Emotion Debate: A Bit of History” by Richard S Lazarus, and chapter 2 “The Philosophy of Cognition and Emotion” by William Lyons, both in T Dalgliesh and M Power (eds) *Handbook of Cognition and Emotion* (John Wiley and Sons Ltd, 1999). See also chapter 8 “Anger” in Dalgliesh and Power (eds) *Cognition and Emotion: From Order to Disorder* (Hove: Psychology Press, 1997) pp 303-343, especially pp 304-310 and pp 323-324. See also De Sousa, chapter 1 “Getting Philosophically Involved With Emotion”, and chapter 7 “The Rationality of Emotion” in *The Rationality of Emotion* (Cambridge and London: The MIT Press, 1987) and Jon Elster, chapter 2 “Emotion” in *Strong Feelings: Emotion, Addiction and Human Behaviour* (Cambridge, Mass: The MIT Press, 1999).

³⁹ “Taking the Heat out of Provocation” (1999) 19 *Oxford Journal of Legal Studies* 481 at 487-8.

⁴⁰ This discomfort can have two consequences: either the process of practical reasoning is polluted by the anger, resulting the conclusion that retribution is the correct response, or practical reasoning functions perfectly well, but the anger tempts the agent to choose in conscious violation of her better judgment. Richard Holton and Stephen Shute discuss the former in “You can’t Lose What You Ain’t Never Had: Self-Control in the Modern Provocation Defence” unpublished manuscript, at 5: see <http://homepages.ed.ac.uk/rholton/pubs.html>. Von Hirsch and Jareborg point out the latter possibility: “The difficulty, it should be noticed, is not purely cognitive. The distinction between permissible and impermissible responses to wrongdoing may be apparent enough” (above n 30 p 250).

feeling of discomfort, is assuaged.⁴¹ Despite this inclination, there are, generally speaking, two ways an agent can exercise self-control so that his anger does not translate into (unjustified) wrongdoing. The first is derived from the nature of the emotion itself. The role of cognition in triggering, and perhaps partially constituting, anger reveals that an important feature of self-control in this context is at the point of cognitive *input*. This is because, for emotionally healthy agents, cognition and affect stand in a proportionate relationship.⁴² Thus the agent who is jostled accidentally and experiences a surge of anger can exercise self-control by focusing on the fact that no deliberate wrong was intended.⁴³ But self-control can also be exercised at the level of behavioural *output*. This is because, as Carroll Izard states, “anger prepares us well for action, but it does not command or compel us to act.”⁴⁴ Such control does not seek to alter the evaluation that triggered the anger, but to ensure that the agent does not allow that anger to translate into the decision to kill or cause serious injury.⁴⁵ For example, it can channel the anger into a non-violent response, such as words. These two forms of self-control are not mutually exclusive, but can reinforce one another. For example, in the case where the agent is jostled accidentally, buying time so as to enable the agent to evaluate whether the conduct was intended to provoke as a matter of input

⁴¹ P S Greenspan *Emotions and Reasons: An Enquiry Into Emotional Justification* (New York: Routledge, 1988), pp 48-55 and pp 159-160. See also her “Subjective Guilt and Responsibility” (1992) 101 *Mind* 287-303 at 288, reprinted in Wallace (ed) *Reason, Emotion and Will* (Aldershot: Ashgate Publishing Company, 1999) pp 475-491, p 476 and her “Unfreedom and Responsibility”, in Schoeman, above n 30, p 67.

⁴² As Michael Moore points out: “Being caused by judgement, such emotions obey laws of proportionality, keeping feeling proportional, both in type and intensity, to its object”, “Choice, Character, and Excuse” in *Placing Blame* (Oxford: Clarendon Press, 1997) p 559, n 29. See also chapter 8 “Anger”, in M Power and T Dagliesh (eds) *Cognition and Emotion: From Order to Disorder* (Hove: Psychology Press, 1997) pp 313-318.

⁴³ As N H Frijda points out in a discussion of the control of emotion: “There is appraisal regulation: Appraisals can be modified within a considerable range by selective attention and self-serving cognitive activities”, see *The Emotions* (Cambridge: Cambridge University Press, 1993) p 401. For evidence of the success of evaluation-alteration reducing anger, see J R Averill *Anger and Aggression: An Essay on Emotion* (New York: Springer-Verlag, 1982) pp 202-203.

⁴⁴ *The Psychology of Emotions* (New York: Plenum Press, 1991) p 248.

⁴⁵ This includes not allowing anger to corrupt judgment, the corruption leading to the conclusion that the provoker deserves what is coming to her: see Holton and Shute, above n 40. Or the agent can control action more directly: as Frijda states: “There is impulse control: Emotional urges can be suppressed so as to disappear from consciousness as well as from behaviour...Overt response can be checked, or attenuated, shaped, or replaced by some other response”, above n 43.

control will require a measure of output control. In turn, the dramatic reduction in, or perhaps extinction of, the agent's anger that flows from her correct assessment facilitates the exercise of output control.

With regard to the difficulties anger creates for self-control, what is key is that when the provoked agent's anger is morally justified, self-control is especially difficult for two reasons. First, input control is not possible: the agent has good reasons for being angry. Secondly, output control involves a moral dilemma, as explained by Andrew Von Hirsch and Nils Jareborg. They point out that, generally speaking, when one is angry, one's moral sense should act as "a bridle to the passions": "Its proper role is to keep anger in check and prevent it from spilling over into action. Ordinarily there is nothing equivocal about that restraining role. If the self-restraint fails, then one's moral inhibitions have not functioned as they should, and that hardly makes one any less to blame."⁴⁶ However, when one is properly angry, '[c]onscience...has a divided role, of encouraging animus against the instigator, prompting one to take certain actions against him, and yet restraining one from making other kinds of responses. This divided role complicates self-restraint in fallible human beings, and our sympathy for the provoked person – our sense of the appropriateness of extenuation – stems therefrom.'⁴⁷ Now, I argued above that the mere fact that the provoked agent's anger is morally justified is insufficient to excuse her for killing in its name. However, when justified anger's implications for self-control are added to the equation, partial exculpation is appropriate. As Gary Watson explains in a criticism of Kahan and Nussbaum, "emotions are both evaluative states and states that potentially interfere with rational control in various ways. It is therefore not surprising that coercive predicaments give rise to defenses expressing both these truths".⁴⁸ The moral defect in the provoked agent's motivation, that is the failure to accord the correct moral weight to the value of the provoker's life and the importance of the lawful resolution of disputes, ceases to be fatal to an excuse in the presence of the challenge to self-control constituted by intense and justified anger. This is because, though we expect agents to refrain from killing purely in the name of private punishment, when self-control is severely challenged by intense and justified anger, we can partially excuse the failure to exercise self-restraint. But both a morally grave provocation and intense anger flowing from that provocation are necessary. Thus the calm agent who kills as a reaction to a wrong, however serious,

⁴⁶ Above n 30, at p 250.

⁴⁷ Ibid.

⁴⁸ "Excusing Addiction" (1999) 18 *Law and Philosophy* 18 589 at 610, n 60.

cannot be excused; similarly, the incensed killer whose anger is generated by petty jealousy cannot be excused, however angry she may be.⁴⁹

The excusatory potential of provocation therefore depends on more than a moral evaluation of the cognitive judgment that triggers, and perhaps partially constitutes, the agent's anger. In light of the challenge to self-control represented by provocation, moral evaluation of the provoked agent, and the potential allocation of an excuse, has a wider focal point, in the form of an assessment of how the agent's capacity for self-control was exercised. This moral evaluation takes into account not only the judgments made by the provoked agent, but the effect of anger on her ability to act on her better judgment. Thus, when the provocation is grave, and the provoked agent experiences intense anger, the conclusion is that self-control has been exercised sufficiently well for the agent to deserve a partial excuse.⁵⁰ Thus the excusatory potential of provocation cannot be confined to an assessment of the judgment at the heart of the provoked agent's anger.

The above criticism of Kahan and Nussbaum has implications for the validity of the shallow capacity theory because, in rejecting possession of self-control as a necessary condition of moral responsibility, it is compelled to conceive of the excuse of provocation in a purely evaluative fashion. It is to be remembered that moral responsibility sets out the necessary and sufficient conditions for entering the process of moral evaluation that may issue in blame, and excuses, full and partial, are concerned with those times when the ascription of blame is not (fully) appropriate as the conclusion of that process. The shallow capacity theory therefore cannot, on the one hand, argue that self-control is irrelevant to moral responsibility whilst, on the other, accept that a moral evaluation of the exercise of this supposedly irrelevant capacity lies at the heart of provocation as excuse. In other words, if the way self-control is exercised lies at the heart of the excuse of provocation, with the result that only those possessed of the capacity can be the subjects of that assessment, then moral responsibility requires possession of such a capacity in this context. After all, as John Gardner often points out correctly, only the morally

⁴⁹ Given Kahan and Nussbaum's purely evaluative approach to provocation, why do they require the agent to be angry at all? Their answer is that anger governs the evaluation made by the agent: "...without anger or rage, the defendant's acts would express nothing more than an inappropriately low valuation of the victim's life", see above n 23, at 315. I surmise that Kahan and Nussbaum believe this is so because an angry killer loses sight of the reasons against killing, whereas a calm killer acts in conscious defiance of those reasons. However, I am far from certain that this is always the case, especially in slow burn cases.

⁵⁰ Horder "Criminal Law: Between Determinism, Liberalism and Criminal Justice" (1996) 49 *Current Legal Problems* 159 at 170.

responsible can make excuses.⁵¹ The result is that a theory of moral responsibility that treats self-control as irrelevant cannot underpin a correct conceptualisation of provocation as excuse.

In contrast, the self-control theory does offer a full account of why provocation can excuse. By acknowledging not only the need for a morally defensible cognitive judgment to trigger (and perhaps partially constitute) the agent's anger, but by also focusing on the difficulties that intense and justified anger creates for the exercise of self-control, the self-control theory completes a picture left incomplete by the shallow capacity theory. In light of this insight, I will now turn my attention to its implications for the criminal law.

The Self-Control Theory and the Criminal Law Partial Defence of Provocation

According to both the shallow and self-control theories, the partial defence of provocation is an excuse, not an exemption.⁵² To this extent, and in contrast to R D Mitchell and B J Mackay's account of the defence,⁵³ the two theories are in agreement. However, as we have just seen, each theory conceptualises the defence's excusatory status differently. The shallow capacity theory, in light of its rejection of the relevance of self-control, conceptualises the defence in a purely evaluative way. It focuses on the moral quality of the cognitive judgment that triggers, and perhaps partially constitutes, the agent's anger. The self-control theory accepts that this is a necessary feature of the defence, but that a full understanding of the defence's excusatory effect depends on integrating the difficulties anger creates for the exercise of self-control. Without those difficulties, an excuse should not be available in light of the provoked agent's failure to accord the correct moral weight to the value of the provoker's life and the lawful resolution of disputes.

⁵¹ This argument is a common theme of all Gardner's works cited in this article.

⁵² The most developed excusatory account of provocation as conceived by the shallow capacity theory can be found in the following articles by John Gardner and Timothy Macklem: "Compassion without Respect: Nine Fallacies in *R v Smith*" [2001] *Criminal Law Review* 623 and "No Provocation without Responsibility" above n 13.

⁵³ The approach of R D Mackay and B J Mitchell is set out in the following articles: "Provoking Diminished Responsibility: Two Pleas Merging into One" [2003] *Criminal Law Review* 745; "Replacing Provocation: More on a Combined Plea" [2004] *Criminal Law Review* 219; and "But is this Provocation? Some Thoughts on the Law Commission's Partial Defences to Homicide" [2005] *Criminal Law Review* 44. Mackay and Mitchell argue that the provocation is really a partial exemption, an approach that I propose to reject in a forthcoming article.

In light of the self-control theory's requirement that the agent experience intense anger, there is much that is correct in the common law requirement, articulated in the first and subjective limb of the defence,⁵⁴ that the defendant experience severe emotional disturbance and a consequent loss of self-control.⁵⁵ The common law is also correct to accommodate the notion of the slow burn: it is not the gap *per se* between the provocative act and killing that should be determinative, but rather the presence of severe emotional disturbance flowing from the provocative act.⁵⁶ Thus the subjective limb plays a necessary role in maintaining provocation's status as an excuse. However, articulations of the subjective limb suffer from a terminological defect that may or may not reflect an underlying philosophical position. The language of *loss* of self-control that features in the case law could be interpreted as implying that self-control is not a capacity, the exercise of which is within the power of the agent, but rather something an agent either has or does not have, with its exercise beyond her control. Thus in *Richens*⁵⁷ the Court talks of the agent being unable to restrain himself. I reject this language and the deterministic thinking that may underpin it, as it is redolent of the hydraulic view of motivation, whereby the agent is at the mercy of her emotions, and whether or not she exercises self-control purely a matter of chance. In light of the notions of input and output control, it is submitted that it is better to say that the provoked agent fails to exercise self-control, rather than loses it, but can be partially excused given the difficulties of so doing when experiencing intense and justified anger. The common law should therefore continue to emphasise the need for a high level of anger, and consequent emotional interference with the exercise of self-control, the result

⁵⁴ The defence is in fact a mixture of the common law and statute, specifically s 3 of the Homicide Act 1957. I will refer to the common law in the text above, as I believe that provocation is best conceived as a common law defence modified by the Homicide Act 1957 s 3.

⁵⁵ See the dictum of Devlin J in *Duffy* [1949] 1 All ER 932 that there must be a "a sudden and temporary loss of self-control, rendering the accused so subject to passion as to make him or her for the moment not master of his mind", approved by the Court of Appeal in *Ibrams* (1981) 74 *Criminal Appeal Reports* and *Thornton* [1992] 1 All ER 306. The Court of Appeal in *Richens* [1993] 4 All ER 877 has stated that there is no need for a complete loss of self-control, in the sense that the defendant does not know what he is doing, merely the inability of the agent to restrain herself. Though I reject the notion of loss of self-control, as I will argue in the text above, these authorities are correct to stress the need for significant emotional disturbance.

⁵⁶ See *Ahluwalia* [1992] 4 All ER 899. Though this is not the place to set out the argument, I believe that many cases of killings by women who are physically abused by husbands/partners should be accommodated by a more sophisticated version of self-defence, as I believe their actions are justified.

⁵⁷ Above n 55.

of which is the agent violating her better judgment, but it should not require that the agent be at the deterministic mercy of her anger.

The recent decisions of *Attorney General for Jersey v Holley*⁵⁸ and *R v Karimi; R v James*⁵⁹ are to be welcomed for rejecting the increasing subjectivisation of the second limb of the defence in light of the characteristics of the provoked agent, a trend exemplified by the decision in *Smith (Morgan)*.⁶⁰ By restoring the objectivity of the second limb, the evaluative element of the defence is preserved through the requirement that the reasonable person would have reacted as the defendant did. This makes clear that the intense anger experienced by the provoked agent must be morally justified.⁶¹ These decisions also serve to ensure that the provocation defence remains an excuse and does not blur into an exemption, the value of preserving this distinction emerging more clearly in the next section of this article.

Thus the common law is more or less theoretically sound, since it requires both emotional disturbance to a significant degree and morally justified anger before the defence is available. In light of this, the recent proposals by the Law Commission for the reform of the defence are of interest, as they appear to be inspired by the purely evaluative view of provocation.⁶² The following are those parts of the Law Commission's proposed reformulation of provocation that are relevant to the present discussion:

- (1) Unlawful homicide that would otherwise be first degree murder should instead be second degree murder if:

⁵⁸ [2005] 3 All ER 371.

⁵⁹ (2006) QB 588.

⁶⁰ [2001] 1 AC 146.

⁶¹ There is much debate as to how to accommodate the various characteristics of the defendant in a manner that respects the evaluative element necessary to provocation's status as an excuse, in particular the workability of Ashworth's distinction between characteristics which bear on the gravity of the provocation and those which bear on the agent's level of self-control ("The Doctrine of Provocation" (1976) 35 *Cambridge Law Journal* 292 at 300). For a discussion, see the works of Mackay and Mitchell above n 53. See also James Chalmers "Merging Provocation and Diminished Responsibility: Some Reasons for Skepticism" [2004] *Criminal Law Review* 198. It is my position that Ashworth's distinction is perfectly workable, and I propose to defend it in another article in which I will reject Mackay and Mitchell's argument that provocation only exists as a partial exemption.

⁶² Above n 24, para 5.11. The effect of this defence is to reduce the Law Commission's proposed category of first degree murder to its proposed category of second degree murder; the Law Commission's definitions of these two categories can be found at paras 1.35 and 1.36.

- (a) the defendant acted in response to:
 - (i) gross provocation (meaning words or conduct or a combination of words and conduct) which caused the defendant to have a justifiable sense of being seriously wronged...
 - (b) [and] a person of the defendant's age and of ordinary temperament, i e, ordinary tolerance and self-restraint, in the circumstances of the defendant might have reacted in the same or a similar way.
- (2) In deciding whether a person of the defendant's age and ordinary temperament, i e ordinary tolerance and self-restraint, in the circumstances of the defendant, might have acted in the same or in a similar way, the court should take into account the defendant's age and all the circumstances of the defendant other than matters whose only relevance to the defendant's conduct is that they bear simply on his or her general capacity for self-control.
- (3) The partial defence should not apply where:
 - (a) The provocation was incited by the defendant for the purpose of providing an excuse to use violence; or
 - (b) the defendant acted in considered desire for revenge.
- (5) A judge should not be required to leave the defence to the jury unless there is evidence on which a reasonable jury, properly directed, could conclude that it might apply.

The requirements for an ordinary temperament and a morally grave provocation in the reformulation are to be welcomed, since, as indicated above and explained more fully in the next section, they help maintain provocation's correct status as an excuse as opposed to an exemption. However, there is no mention of the effect of anger on self-control. Indeed there is no mention of the agent's emotional state at all. This is a deliberate omission by the Law Commission, as it believes the subjective limb of the defence to be unnecessary and undesirable.⁶³ However, the omission is a serious one, because, as we have seen, the effect of anger on self-control is

⁶³ Ibid, paras 5.17 to 5.32.

key to the defence's excusatory status.⁶⁴ To object to this requirement is therefore to deny the inherent nature of the defence. Nevertheless, the Law Commission advances three arguments against the subjective limb; these arguments, however, can be countered if the self-control theory's conceptualisation of the subjective limb is adopted.⁶⁵

First, it believes that the subjective limb suffers from one particular ambiguity: does it mean that the defendant was unable to exercise self-control, or simply failed to exercise self-control? Furthermore, it believes the first question cannot be answered, even by medical science.⁶⁶ However the self-control theory, rejecting as it does the hydraulic view of human motivation, resolves this ambiguity in favour of the defendant having failed to exercise self-control, with an excuse only being available when this failure can be explained in terms of the difficulties created by intense and justified anger. In so doing, it also avoids the supposed impossibility of ascertaining whether the defendant was unable to restrain himself. Second, the Law Commission argues that the current defence favours male reactions over those of women, making it more difficult for women to satisfy the defence.⁶⁷ But severe anger is not the preserve of men, and the defence is designed for people who kill whilst experiencing such anger. There is a need for a defence for women who kill oppressors, but provocation is not that defence, and should not be altered beyond recognition to meet that need.⁶⁸ Third, the general tenor of the Law Commission's rejection of the subjective limb is that it is too vague, and

⁶⁴ Thus Mackay and Mitchell are correct to point out that emotional disturbance is a necessary feature of provocation, see "But is this Provocation?" above n 53, especially at 46-47.

⁶⁵ The Law Commission quotes with approval the arguments that it set out in Law Commission *Partial Defences to Murder* Law Com No 290 (London: TSO, 2004) ('Law Com 2004'), paras 3.28 to 3.30 and 3.135 to 3.137, see above n 24, para 5.2 and n 13 of Part 5. The argument in the text above is therefore a response to the arguments set out in both the 2004 and 2006 Reports.

⁶⁶ Law Com 2004, *ibid*, para 3.28.

⁶⁷ Law Com 2006, above n 24, para 5.18.

⁶⁸ Thus I accept that there is a need for a partial defence based on a killer's fear of serious violence: see s1(a)(ii) and (iii) of the Law Commission's reformulation, Law Com 2006, above n 24, para 5.11. Such a defence, based on forward looking reasons, could, as a result, be purely evaluative (on the significance of forward looking reasons, see my critique in the text above of Kahan and Nussbaum's evaluative conception of provocation). One would imagine that the fear of serious violence would have to be based on reasonable grounds. However, it is arguable that such a defence does not go far enough for certain types of victim, such a physically abused women, and that there is a need for a justificatory defence for such circumstances resulting in a full acquittal. Such a defence would require a more relaxed approach to the necessity requirement than currently available for self-defence.

therefore unworkable.⁶⁹ In this regard, it argues that the acceptance of the notion of the slow burn, designed to reflect how anger may function where women are concerned, only adds to this lack of clarity.⁷⁰ But the notion of extreme anger as required by the self-control theory is not difficult to define: it would not lead, as argued by the Law Commission, to intense legal scrutiny.⁷¹ Nor is it beyond the experience of jurors: they are more than capable of asking, and answering, the question: did this person kill whilst experiencing extreme anger? It is no more difficult than many other questions of law that they are asked to understand and apply.⁷²

The Law Commission would appear to believe that the benefits of the subjective limb are retained with greater clarity by its prohibitions on considered revenge and engineered provocation.⁷³ I am not so sure. It seems to me that the concept of considered revenge is much more likely to engender complex legal debate than extreme anger, involving finely tuned argument as to what constitutes revenge, and then what is a considered version thereof. The prohibition on engineered provocation is in fact not relevant to the merit or otherwise of a subjective limb, but simply a moral bar to the use of the defence easily bolted on to the subjective limb. Furthermore, it is vital that the defendant acts against her better judgment in killing, a requirement presently captured by the subjective limb, and the Law Commission's bar on considered revenge does not guarantee that it is in place. The fact that a person of ordinary temperament might have reacted in the same way, when combined with no mention of anger's effect on self-control, makes the reformulation consistent with a spontaneous response that is nevertheless in accordance with the agent's values. Such a defendant, acting in accordance with her better judgment, should not benefit from the defence, yet the substantive content of the Law Commission's reformulation would make it available to such a defendant. Perhaps the Law Commission imagines that the judge would prevent such cases reaching the jury under (5) of the reformulation, but if so, upon what basis would he or she do so? If it is because the judge feels that an agent should not benefit from the defence without losing self-control, then the reformulation will have failed to remove the subjective limb from the defence.

⁶⁹ Law Com 2004, above n 65, para 3.30, Law Com 2006, above n 24, paras 5.17 and 5.19.

⁷⁰ Law Com 2004, *ibid*, para 3.29.

⁷¹ Law Com 2006, above n 24, para 5.22, when discussing the notion of extreme emotional disturbance.

⁷² In this respect, the Law Commission recognises that the notion of extreme emotional disturbance may indeed be workable: see Law Com 2006, above n 24, para 5.22, and may even be required to be incorporated into the defence: para 5.32 (2).

⁷³ Law Com 2006, *ibid*, para 5.20.

It is worth noting that the Law Commission accepts that 1(a)(i) of its reformulation may also require ‘a spontaneous display of anger’, but only suggests this to avoid certain premeditated killings from benefiting from the defence, as opposed to incorporating the requirement that the agent be acting against her better judgment.⁷⁴ However, acting against better judgment is a key element of the excuse. The notion of premeditation is relevant to this question, in the sense that premeditated killings will almost always be in accordance with the agent’s better judgment, but some killings that are not premeditated will also be in accordance with the agent’s better judgment, and these should also be precluded from the defence. The prohibition on considered revenge, married with a requirement of spontaneous anger, does not therefore suffice to exclude such unmeritorious cases.

To conclude, so long as the subjective limb is gently re-conceptualised in accordance with the self-control theory, it would be a mistake for the criminal law to abandon it as the Law Commission invites it to do. The notions of anger and failing to exercise self-control, key elements of provocation’s excusatory power, should be explicit and prominent features of the defence.

I have argued that because, under certain circumstances, a provoked killer exercises her capacity for self-control with a certain level of normative competence, she should be partially *excused* for her wrongdoing. This is because the provoked agent is not seeking to exempt herself from the moral standards and modes of assessment that govern the allocation of excuses, but rather to submit herself to those standards and modes of assessment, and show how she more or less met them. More specifically, she is not claiming that her capacity for self-control was lost, impaired or underdeveloped, but that it was intact and exercised to a certain standard.⁷⁵ However, on other occasions, the agent’s capacity for self-control will be adversely affected through no fault of her own. On such occasions, full or partial exemption from the standards inherent in the allocation of excuses may be appropriate. This brings me to the second flaw in the shallow capacity theory.

⁷⁴ Law Com 2006, *ibid*, para 5.30

⁷⁵ Sometimes commentators talk of self-control being impaired by anger but this is based on confusion created by anger’s internal quality; however, despite this internal quality, anger *per se* does not impair self-control because it is exactly the type of phenomenon, along with, for example, incontinent desires, that self-control was designed to deal with. To say otherwise is akin to claiming that a person’s strength is impaired by the weight she is lifting. I reject the notion that self-control is a finkish disposition, disabled by anger, the very stimulus it is meant to control: see Holton and Shute, above n 40, n 30 and accompanying text.

THE SHALLOW CAPACITY THEORY'S SECOND FLAW: AN INCOMPLETE ACCOUNT OF EXEMPTIONS

We have seen how the shallow capacity theory gives a convincing account of why automatism, insanity and infancy exempt, because such phenomena mean that the agent is either not the author of her wrongdoing or fundamentally incapable of rational action. However, its second flaw is that it cannot account for the exempting effect of other phenomena that should also exempt; these are involuntary intoxication by alcohol, drug addiction and phobias. In this section of the article, I will expose this flaw by arguing that the correct way of conceiving these phenomena is by showing how they can create such severe difficulties for self-control that the capacity should be considered impaired, thereby exempting the agent from a moral evaluation of how it has been exercised. Given its rejection of the relevance of self-control, this argument is not available to the shallow capacity theory. Based on this insight, I will then argue that there is a need in the criminal law for a distinct defence based on the notion of radical impairment of the agent's capacity for self-control.

Exposing the Flaw in the Shallow Capacity Theory's Account of Exemptions

As I have just stated, the shallow capacity theory cannot account for the exempting effect of involuntary intoxication, drug addiction or phobias by means of any argument that refers to the agent's capacity for self-control. If it is to give these phenomena exempting status, it must do so in one of two other ways. Either it must show that their presence means the agent is incapable of behaving intelligibly, and so account for them in the way it accounts for infancy or insanity. Alternatively, it must argue that their presence means the agent is not the author of the wrongdoing, and so account for them in the way that it accounts for wrongdoing resulting from a seizure or a reflex. If it can do neither, and I will explain that it cannot, then it must accept that the agent is fully morally responsible, which I will argue is wrong. Alternatively, it may try to account for the exculpatory pull of such examples in the context of excuses, but I will show that this attempt also fails.

Let me begin with involuntary intoxication by alcohol. Are wrongdoings performed whilst drunk, even very drunk, unintelligible? Except on those rare occasions when alcohol consumption is such that the agent is rendered insane or an automaton, actions under the influence of alcohol are perfectly intelligible, even if often self-destructive, extreme or cruel. This can be illustrated by the case of *Kingston*.⁷⁶ Though this case concerns drugs other

⁷⁶ [1994] 3 All ER 353.

than alcohol, they were disinhibitors, and therefore can be considered the same as alcohol for the purposes of this discussion.⁷⁷ Kingston was a homosexual man with paedophilic tendencies that he had, apparently, hitherto refrained from indulging. Invited to a flat in the belief that it was to discuss business matters, he was given coffee laced, without his knowledge, with soporific drugs.⁷⁸ He was then taken into a bedroom where he found a 15 year-old-boy lying unconscious on the bed. He abused the boy sexually. Despite the disinhibiting effect of the drugs, Kingston's behaviour was intelligible, though morally reprehensible: he was confronted with an object of sexual attraction, and he took advantage of that opportunity. Perhaps, then, he was not the author of his actions? This argument would not appear available to the shallow capacity theory, since its refusal to compartmentalise the agent at the psychological level means that Kingston is identified with the values expressed in the act of abuse. And the argument that the drugs had severed the link between Kingston and his conduct seems equally problematic; at best, it can be said the drugs distorted Kingston's values. But were they distorted? In this regard, the adage *in vino veritas* springs to mind: Kingston may sincerely believe that he should not indulge his desire for under-age boys, perhaps only out of fear of punishment, but he will certainly value so doing at some level. Kingston's agency as conceived by the shallow capacity theory therefore appears to have been expressed in his act of abuse, even though he failed to act in accordance with his better judgment. Perhaps then he can be excused? However, an evaluation of his motivation reveals that an excuse is not available: he has acted for bad reasons and failed to accord the correct moral weight to his victim's physical integrity and emotional well-being. He is therefore fully culpable according to the shallow capacity theory. Yet I resist this conclusion: Kingston's capacity for self-control, his capacity to act for the reasons he deemed best, was severely compromised by his unknowing consumption of disinhibiting drugs. This factor should result in, at the very least, a partial exemption. I will explain why in greater detail below.

The shallow capacity theory similarly cannot account for the impact of drug addiction on moral responsibility.⁷⁹ The wrongdoing in question is consumption of the relevant drug(s) or wrongdoing designed to enable

⁷⁷ For a full discussion of this case, see G R Sullivan "Making Excuses" in A P Simester and A T H Smith (eds) *Harm and Culpability* above n 8, pp 131-152.

⁷⁸ "There was evidence that the defendant had been slipped three kinds of soporific drugs with properties which would affect judgment, consciousness, and memory": see Sullivan, *ibid* p 131, n 3.

⁷⁹ I am leaving to one side the important question of the drug addict's moral responsibility for her wrongdoing that is derived from her original decision to consume the drug(s). For a discussion of this question, see S J Morse "Hooked on Hype: Addiction and Responsibility" (1999) 18 *Law and Philosophy* 3 at 24-27.

consumption, such as burglary or theft. The addict's decision to commit such wrongs is intelligible.⁸⁰ Where consumption is concerned, it expresses an intelligible decision to be free of, or satiate, the intense desire/craving;⁸¹ if it is related wrongdoing, it expresses that value and an intelligible decision to violate the rights of others. From the point of view of authorship, it makes no difference to the shallow capacity theory that the addict disapproves of those wrongful choices motivated by her addiction, because the shallow capacity theory identifies the agent with the values expressed by those choices. It is accepted that, if the behavioural implications of drug addiction are irresistible,⁸² the wrongdoing does not reflect the addict's values, since such wrongdoing is nothing more than the direct expression of appetite free of value. In such cases, the intelligibility of the wrongdoing is not in question, since the addict is no longer acting for reasons, and, as a result, no longer the author of her wrongdoing. But such extreme cases, where the addict is akin to an automaton, must be rare. Thus the addict, except for rare occasions, is fully morally responsible according to the shallow capacity theory. Since she has also acted for bad reasons, and perhaps also failed to accord the correct moral weight to the interests of others, she cannot claim an excuse either. The result is that she is fully culpable according to the shallow capacity theory. Yet, once more, I resist this conclusion. The powerful cognitive focus that addiction brings on the pleasures of consumption, a focus that makes acting on better judgment very difficult, should be acknowledged as a form of exculpation, in the form of a partial, and perhaps on some occasions, a full exemption. I will also explain why below.

It is accepted that with phobias, the emotion of fear could be said to be unintelligible, in the sense that its object is not a rational source of fear, or, if that object is a rational source of fear, the intensity of the emotion is disproportionate to that object. But wrongdoing flowing from the experience of that fear,⁸³ if not the phobia, is intelligible, in the sense that the agent expresses an intelligible decision not to experience the symptoms of the phobia. With regard to authorship, the phobic agent's disapproval of her failure to overcome the behavioural implications of the phobia changes

⁸⁰ I am concerned with wrongdoing committed whilst the addict is experiencing desires/cravings. Wrongdoing at other times is a question of excuse.

⁸¹ For a sophisticated discussion of how to conceptualise addiction, see J Orford "Addiction as Excessive Appetite" (2001) *96 Addiction* 15.

⁸² For an explanation of why addictive desires can be resisted, see Wallace "Addiction as Defect of the Will: Some Philosophical Reflections" above n 2 at 621-624. For a discussion of those extreme cases where the desire might be irresistible, see J Elster above n 38, p 138.

⁸³ For example, an acrophobic father's failure to rescue his child from the edge of a cliff.

nothing where the shallow capacity theory is concerned, because the shallow capacity theory nevertheless identifies the agent with the values expressed by her choice to succumb. It is accepted that, if the behavioural implications of the phobia were irresistible, it could not be said that the wrongdoing reflects the agent's values, since the intensity of the emotion, combined with its shallow cognitive input (e.g. 'this is a spider'), has nothing to do with the holding of values. But such extreme cases, where the emotion has direct and irresistible access to the agent's conduct, must be rare. Thus, except for rare occasions, the phobic agent is fully morally responsible according to the shallow capacity theory.

Is it possible to account for the exculpatory effect of phobias within the domain of excuses, through an assessment of the values expressed in the agent's wrongdoing? Such an approach would enable the shallow capacity theory to account for the exculpatory pull of such cases. Eugene Schlossberger has attempted to do so.⁸⁴ Treating the phobic as morally responsible, Schlossberger argues that the phobia means that *what* the agent is morally responsible *for* is less blameworthy. This is so, Schlossberger argues, because of the positive moral features of the intelligible reasons for which the phobic agent acts, as is the case with certain instances of duress. His example is that of a snake phobic who will not pick up a friend's pet garter snake as it escapes past his feet. Schlossberger argues that for the non-phobic, picking up the snake is a minor inconvenience, whereas for the phobic, so doing may involve great personal cost in overcoming the aversion. Exculpation is therefore appropriate. This example is convincing because Schlossberger chooses it well: arguably, the discomfort of overcoming the aversion can be weighed morally against the loss of the snake. However, if one imagines that the snake phobic is an immunised mother who fails to rescue her non-immunised child from a room containing a poisonous snake, the same conclusion cannot be reached. The blameworthiness of the failure to rescue the child is undiminished because no amount of discomfort has moral weight compared to the interests of the child. As we have seen, the same goes for the other examples above: the intoxicated child molester and the drug addict act for bad, albeit intelligible, reasons. Excuses are not available to such agents. If exculpation is to be found, it is in the context of exemptions, or not at all. And the shallow capacity theory's exclusive focus on authorship and rationality means exemptions are not available.

Unlike the shallow capacity theory, a theory of moral responsibility that acknowledges the necessary role of the capacity for self-control can cope with the exempting effect of the above examples in a way that chimes with intuition. Accepting that the involuntarily drunk, drug addicted and phobic

⁸⁴ Above n 3, at p 153.

agent is normally the intelligible author of her actions, such a theory nevertheless exculpates in the form of a full or partial exemption when the agent's capacity for self-control is sufficiently compromised by the effects of alcohol, addiction or phobia. The capacity for self-control is compromised in such cases because involuntarily intoxicated and addicted and phobic agents are not, for various reasons, fully cognitively or emotionally healthy; specifically, these conditions (temporary or otherwise) have direct implications for the psychological processes that govern or underpin the exercise of self-control. Such implications take different forms: where alcohol is concerned it consists of the weakening of those mental faculties of inhibition,⁸⁵ where addiction is concerned, it consists of the powerful cognitive focus on consumption generated by the addiction,⁸⁶ where phobia is concerned, it consists of intense and irrational emotion generating a powerful avoidance reaction.⁸⁷ Such effects should not be confused with the fundamental incapacity for intelligible action of the infant or the paranoid schizophrenic: as argued above, the involuntarily drunk, addicted and phobic agent's behaviour remains rational, that is based on an intelligible interpretation of, and response to, her situation. In these cases, the problem is elsewhere, in the form of specific difficulties for the successful exercise of the agent's capacity for self-control. Now these difficulties may also be a form of irrationality, but if so, they are of a particular kind, fundamentally different to the core irrationality of the insane or core non-rationality of the infant. In turn, their presence frees the person from the standards of self-control we expect from the sober, addiction free and emotionally healthy. However, whether, and the degree to which, exemption is appropriate in such cases will depend on the intensity or degree of the compromising or burdening effects of the potentially exempting factor.⁸⁸

Cases of involuntary intoxication, drug addiction and phobia may be contrasted with provocation where the anger experienced as a result of the provocation is appropriate and proportionate, and the agent's capacity for self-control, specifically her ability to exercise input control over the emotion and output control over behaviour, is intact.⁸⁹ This is why the Law Commission's

⁸⁵ See, for example, Mark T Fillimore and Jessica Weafer "Alcohol Impairment of Behavior in Men and Women" (2004) *99 Addiction* 1237.

⁸⁶ Morse, above n 79.

⁸⁷ Susan Nolen-Hoeksema *Abnormal Psychology* (McGraw-Hill, 3rd edn, 2004) p 170.

⁸⁸ Wallace offers an explanation of how to conceptualise that intensity in "Addiction as Defect of the Will: Some Philosophical Reflections", above n 2. So does Morse, see above n 79.

⁸⁹ This may be contrasted with those suffering from phobias, where input control over the emotion of fear has been more or less lost.

proposals for reform of the defence of provocation are correct to require that the agent be of “ordinary temperament”.⁹⁰ But the potential for excuse in instances of provocation is merely a feature of a more general truth about the manner in which moral judgment and blame function in the context of any kind of test of an agent’s capacity for self-control. This is that we expect cognitively and emotionally healthy agents to exercise self-control in the face of a range of phenomena that tempt the agent to act against her better judgment, and we judge those who have failed to exercise self-control in the expected manner more or less harshly. In turn, we exempt from that expectation those we know to be insufficiently emotionally and cognitively healthy, as can be the case with the involuntarily drunk, addicted and phobic agent.

The overarching argument for the above approach to moral responsibility lies in a vision that treats agents as fully morally responsible only when their capacity to control their actions in light of their better judgment is more or less intact. This vision conceives of moral responsibility as requiring agents who are more than passive, albeit intelligible, reactors to the world, but meaningful and creative contributors to the world. The existence of the capacity to evaluate normatively and develop conceptions of the good, and then to control conduct in light of those conceptions, lies at the heart of that type of meaningful and creative interaction. In turn, when the agent’s capacity for self-control, and hence her capacity for such interaction, is severely compromised, as it can be in the instances outlined above, then the agent’s moral responsibility has also been compromised.

Unlike the shallow capacity theory, therefore, a theory that recognises that possession of the capacity for self-control is a necessary condition of moral responsibility does not limit the notion of exemption to denials of authorship or core rational capacity: it includes a third kind of exemption turning on the compromise of the agent’s capacity for self-control. I will now turn my attention to the implications of this argument for the criminal law.

Impaired Self-Control: The Need for a New Criminal Law Defence

English criminal law does not include a distinct defence based on the impairment of self-control. Involuntary intoxication, drug addiction and phobias are acknowledged only to the extent that they negate *mens rea* or fall

⁹⁰ Agents who have specific and medically recognized difficulties with anger, for example those suffering from borderline personality disorder, may also be considered as having an impaired capacity for self-control. They are therefore suitable candidates for the diminished responsibility defence, as well as a more general defence of radically impaired self-control, the need for which I describe in the next sub-section of this article.

within the diminished responsibility defence. However, involuntary intoxication and the like will not always negate *mens rea*: the case of *Kingston*,⁹¹ discussed above, is an example. Furthermore, diminished responsibility is a defence confined to murder.⁹² As a result, criminal law commentators who recognise that imposing criminal liability in such cases is problematic have argued that there is a gap in the lexicon of criminal law defences.⁹³ The question is how should such a new defence be conceptualised, and hence accommodated, by the criminal law? My aim in this part of the article is to show that the self-control theory offers a way to conceptualise such defences that is superior to the shallow capacity theory. Its superiority is due to the shallow capacity theory's insistence, once the agent is author of her wrongdoing, that all denials of moral responsibility take the form of denials, at the time of wrongdoing, of the agent's core capacity for rational or intelligible action. This is too narrow an approach. This excessive narrowness can be illustrated with Gardner's account of the partial defence of diminished responsibility, and how he distinguishes that defence from the partial defence of provocation, because his argument is informed by the shallow capacity theory.⁹⁴

Gardner argues that all excuses are cases of what he terms "indirect rational explanation."⁹⁵ This means the agent does not directly justify her wrongdoing, which remains irrational, but rather the beliefs, emotions or desires that led to it. Thus provocation excuses when the wrongdoer can point to intelligible and good reasons for her anger. In keeping with the shallow capacity theory, it is a purely evaluative approach that focuses on the moral quality of the agent's rational reasons for feeling as she does. In turn, according to Gardner, the way that provocation differs from diminished

⁹¹ Above n 76.

⁹² Homicide Act 1957 s 2.

⁹³ Bob Sullivan "Making Excuses" in *Harm and Culpability* above n 8, pp 131-152 and Victor Tadros "The Characters of Excuse" (2001) 21 *Oxford Journal of Legal Studies* 495-519. These commentators argue that such cases should be based on the notion of "lapse of character". I believe the character theory of moral and criminal responsibility is flawed, but this is not the place to explore its defects. Lacey, who has also acknowledged the need for such defences, calls such cases "situations of blocked evaluation." Lacey's approach appears to have something in common with the self-control theory, because she expresses a preference for treating such cases as involving the removal of, or interference with, "the normal reasoning processes", see above n 11 at p 121.

⁹⁴ It is worth noting that Gardner is talking about diminished responsibility where agent has been provoked, as opposed to a general analysis of the defence of diminished responsibility.

⁹⁵ See "The Mark of Responsibility" above n 13, at 158.

responsibility is that, with diminished responsibility, there is no rational explanation for the agent's anger, as well as the actions that flow from it. The rationality defect goes "all the way down".⁹⁶ In other words, the agent is fundamentally incapable of intelligible action, and as a result a partial exemption, and hence partial defence, is appropriate. With provocation, however, the rationality defect is confined to the action itself, but not the anger upon which it is based, which remains, at the very least, partially justified.⁹⁷ The fact of, at the very least, partial justification means that the agent is not denying her moral responsibility, but actually asserting it.⁹⁸

At first blush, it is hard to see how Gardner's approach is compatible with the notion of partial responsibility. If the rationality defect with diminished responsibility goes all the way down, why isn't a full exemption, and hence a full defence, appropriate? This objection is perhaps dealt with by treating irrationality as a matter of degree, as is the case with children as they progress toward adulthood, and treating the agent's anger as only partially irrational. However this is an awkward solution and, in any event, not one advanced by Gardner, who always talks of the required irrationality in absolute terms. Furthermore, this response only deals with the notion of partial responsibility within Gardner's rationale, and that rationale is incomplete. To explain why, we must turn our attention to the phenomena of the previous subsection.

Where involuntary intoxication and drug addiction are concerned, as I have argued above, the agent's assessment of, and response to, her situation will often remain rational. For such agents, the problem is not the fundamental incapacity for intelligible action that Gardner describes, but acute difficulties in controlling behaviour in accordance with better judgment. Thus Gardner's rationale cannot account for the exempting power of these phenomena. The phobic wrongdoer falls more plausibly into Gardner's rationale because, arguably, the rationality defect does go all the way down in such cases: an irrational action out of an irrational fear, that is to say flight or avoidance when there is no real danger. But this congruence is superficial, because, as we have seen, the decision not to experience the effects of the phobia is rational, even if the fear is not. This can be distinguished from anger in the context of diminished responsibility, where the action that flows from the anger is contaminated by the irrationality of the emotion itself, hence the plausibility of Gardner's approach in such cases. Thus Gardner's rationale also fails to account for the exempting effect of phobias. Furthermore, and equally importantly, Gardner's approach arguably will exempt on certain occasions when exemption is not appropriate. Even if the emotion, be it fear

⁹⁶ "No Provocation without Responsibility...", above n 13, at 214-215.

⁹⁷ *Ibid*, at 214.

⁹⁸ *Ibid*, at 215.

or anger, is irrational and in some sense leads to irrational action, no exemption, and hence no defence, should be available if the anger or the symptoms of phobia are nevertheless relatively mild. This is because, if they are mild, the agent is expected to exercise self-control over her conduct, and thus avoid any wrongdoing.

Thus the shallow capacity theory, with its exclusive focus on the agent's core capacity for rational action, is not equipped to deal with why and how we exempt in cases of involuntary intoxication, drug addiction and phobia. This is because whether an exemption, full or partial, should be allocated to the involuntarily drunk, addicted or phobic agent cannot be decided in terms of her capacity for rational action, but must be ascertained in terms of how difficult it is for such an agent to exercise self-control successfully. Thus, for example, if the agent is not that intoxicated, or the drug addict's cognitive focus on consumption is not that intense, no exemption should be available. The self-control theory therefore offers a model for understanding when a partial or even a full exemption, and hence defence, is appropriate, and when it is not, in the form of the intensity or severity of the factor compromising the exercise of self-control. The shallow capacity theory, given its rejection of the relevance of self-control, has no sense of this 'difficulty', with the notable consequence that it runs the risk of exempting when it should not simply because irrational behaviour is linked in some way to an irrational emotion.

Thus the criminal law should create a new defence based on the notion of radically impaired self-control. I do not propose to analyse the exact form such a defence should take, or whether it should result in a complete acquittal or a special verdict of some description, the latter perhaps leading to civil commitment procedures where there is a danger to the public.⁹⁹ I also acknowledge that the administration of such a defence raises issues over the allocation of the burden of proof, and also important forensic questions. Suffice to say for present purposes is that, if the criminal law wishes to reflect accurately the moral culpability of agents, the self-control theory reveals that there is a need in the criminal law for a defence based on the notion of radical impairment of the agent's capacity for self-control.

CONCLUSION

In conclusion, I would like to make some general comments about the fundamental nature of the conflict between the shallow capacity theory and the self-control theory. This conflict, I believe, is a conflict over how

⁹⁹ It is worth noting that the defence would resemble the Law Commission's recent proposal for the amendment of the diminished responsibility defence, though it would have wider application: see above n 24, para 5.112, specifically (a)(iii).

profound the judgement of blame should be. The self-control theory accepts that blame for wrongdoing is, as Wallace puts it, “a form of deep moral assessment that goes beyond mere evaluative description of what the agent has done.”¹⁰⁰ The problem with the shallow capacity theory is that it does not look beyond the agent’s raw capacity for intelligible motivation, with the allocation of excuses dependent on the moral quality of that motivation. The problem is that, if blame is to have depth, it must be based on more than a capacity for intelligibility and consequent pejorative description of the values expressed by the agent’s wrongdoing: it must also pre-suppose an agent who is, in some sense, *in control* of her conduct in a way that is meaningful. And if this is so, moral responsibility requires more than a mere capacity to conduct oneself in an intelligible fashion.¹⁰¹ This something more takes the form of the capacity for self-control. It is this capacity that gives human agents a unique form of dignity. It is from this dignity that blame for wrongdoing derives the depth that I believe is central to its identity as a form of moral judgment. But I appreciate that this requires much more justification.

¹⁰⁰ *Responsibility and The Moral Sentiments* (Cambridge, Mass: Harvard University Press, 1996) p 75.

¹⁰¹ The idea that moral responsibility requires more than exhibiting phenomena capable of being labelled as bad is to be found in work by the following authors: Wallace *ibid*; S Wolf *Freedom Within Reason* (Oxford: Oxford University Press, 1990); M J Zimmerman “Responsibility Regarding the Unthinkable” (1995) 20 *Midwest Studies in Philosophy* 204-223 at 213-215; J A Montmarquet *Epistemic Virtue and Doxastic Responsibility* (Lanham, Md: Rowman & Littlefield, 1993) pp 68-69; and J Sabini and M Silver, “Emotions, Responsibility and Character”, in Schoeman (ed) *Responsibility, Character and the Emotions: New Essays in Moral Psychology* above n 30, pp 165-175. Zimmerman has pointed out in a criticism of Schlossberger that there is a difference between moral responsibility and exhibiting phenomena capable of moral evaluation (p 213).

The principal aim of this paper has been to outline two fundamental flaws with a theory that treats possession of the capacity for self-control as irrelevant to moral responsibility for wrongdoing, and thereby demonstrate the superiority of a theory that recognises that possession of such a capacity is a necessary condition of moral responsibility. A full exploration of the exact nature of this capacity for self-control, and a full defence of why its possession is a necessary condition of moral responsibility for wrongdoing, is a project for another time.