

© Editions l'Harmattan.

URL:

<http://www.editions-harmattan.fr/index.asp?navig=catalogue&obj=livre&no=44772>

A SHORT GUIDE TO THE ANGLO-AMERICAN DEBATE ON CRIMINAL RESPONSIBILITY

James Slater

Senior Lecturer in Law at the University of Buckingham

INTRODUCTION

Examination of the rules and principles of Anglo-American criminal law reveals that there is often more to criminal liability than the presence of criminal wrongdoing per se, understood as certain instances of harm risking or harm causing behaviour.¹ The frequent requirement for mental states such as intention, recklessness and belief make a conviction on certain occasions not only dependent on wrongdoing, but also the presence of a morally culpable state of mind; when available, the defences of loss of control/provocation and duress, with their reference to the conditions and emotional states under which intentional wrongdoing occurs, accept that, though wrongdoing is in place, exculpation in some form is nevertheless appropriate; the focus on mental health exhibited by the defences of insanity and diminished responsibility, and on age by the defence of infancy, reflects a concern not with the wrong committed, but with the defendant's mental faculties and development at the time of wrongdoing. All of these elements of Anglo-American criminal law exhibit a common feature: a concern not so much with what the accused has done, but with whether, how and why he should be held to account for having done it. In seeking theoretical understanding of such elements, the theorist is not engaged in understanding the grounds upon which certain behaviour is or should be criminalised, involving matters such as the role of the harm principle or the legitimate reach of inchoate liability; his concern is with the grounds of liability once criminal wrongdoing is in place, and it is those grounds that constitute the notion of *criminal responsibility*.

Criminal responsibility has generated considerable debate in Anglo-American criminal theory.² Contributions to this debate have sought to describe the theoretical underpinnings of criminal responsibility and to offer a blueprint for its normative development. In this debate, two theories vie for descriptive and normative dominance:

* I would like to thank Jeremy Horder, Carol Brennan and March Stauch for their very helpful comments on earlier drafts of this article. Needless to say, any errors remain my own.

¹ For the sake of completeness it should be acknowledged that some criminal wrongs take the form of omissions in the face of a duty to act, as opposed to harm risking or causing actions. Criminal wrongdoing also embraces inchoate and secondary liability: arguably these are also forms of harm risking or harm causing action.

² Rather than listing all of those who have contributed to the debate here, I will instead reference the work of contributors at those points in the text when I refer to their ideas.

character and capacity.³ This article is conceived as a guide through this debate for those possessing some general familiarity with Anglo-American criminal law,⁴ but who are not immersed in the debate. The aims of this article are threefold: to give an overview of the various arguments that inform the debate, supply theoretical perspective on the nature of the conflict between the rival theories and, finally, to propose an overall winner.

PLACING THE DEBATE IN CONTEXT

Most theoreticians who have entered the debate on criminal responsibility treat the criminal law's rules and principles of criminal responsibility as concerned with the question of moral culpability. There are two alternative justifications for conceptualising them in this way. For those theorists of what may be termed the 'pure' school, the justification is borne of their vision of the criminal law as exclusively concerned with punishing the morally culpable for moral wrongdoing: given that, for such theorists, the exclusive goal of the criminal law is to track moral culpability, the rules of criminal responsibility should be conceptualised in light of that goal.⁵ Other theorists of what is often known as the 'mixed' school do not see the criminal law as exclusively concerned with punishing the morally culpable but rather draw a distinction between a utilitarian general justifying aim of punishment, normally in the form of deterring and preventing socially harmful behaviour, and a principle of individual distribution of punishment based on moral culpability.⁶ For such theorists moral culpability functions as a constraint on the pursuit of a utilitarian goal. The principle of individual distribution means mixed theorists believe, in common with pure theorists, that the criminal law's rules and principles of criminal responsibility should be understood as ensuring the moral culpability of the criminal wrongdoer.⁷

³ For those already familiar with the debate on criminal responsibility who are wondering why no mention has been made of choice theory, please note that I deal with choice theory as a species of capacity theory.

⁴ From now on, references to the criminal law should be taken as references to Anglo-American criminal law.

⁵ The leading work setting out such a view is Michael Moore's *Placing Blame* (Oxford: Oxford University Press, 1997). Moore sees the criminal law as a "functional kind whose function is to attain retributive justice" by "punish[ing] all and only those who are morally culpable in the doing of some morally wrongful action." p 33 and p 35. This vision of criminal responsibility underpins a retributivist theory of punishment, whereby the convicted party, in view of his culpability, deserves punishment in some form.

⁶ An influential formulation of the mixed theory is to be found in H L A Hart's *Punishment and Responsibility: Essays in the Philosophy of Law* 2nd ed (Oxford: Oxford University Press, 2008), especially pp 8-13. For a powerful critique of mixed theories see Nicola Lacey *State Punishment: Political Principles and Community Values* (London: Routledge, 1988), pp 46-57. At first Hart's principle of distribution was arguably inspired by utilitarian calculations, creating a system of choice that enabled people to avoid the ambit of criminal regulation and thereby maximise their freedom: see 'Legal Responsibility and Excuses' in *Punishment and Responsibility*, p 28. His second and later justification was arguably based on moral considerations, see 'Murder and the Principles of Punishment: England and the United States' in *Punishment and Responsibility*, p 80: there are constraints "which civilised moral thought places on the pursuit of the utilitarian goal by the demand that punishment should not be applied to the innocent." In any event it should be noted that the mixed theory has several variants and the general description in the text above is a simplification. There are also versions of the mixed theory that reverse the relationship between desert and utilitarian rationale, treating the latter as underpinning distribution: once again, see Lacey *State Punishment* pp 53-56.

⁷ For a general discussion of the relationship between punishment, moral desert and criminal responsibility, see Lloyd Weinrib 'Desert, Punishment, and Criminal Responsibility' (1986) 49 *Law and Contemporary Problems* 47, especially at 47-53.

Both character and capacity theorists have advanced their theories as the best theories of moral culpability in the service of either the pure or mixed approaches outlined above. By best, I mean that those theorists believe their theories give both an accurate descriptive account of the existing rules and principles of criminal responsibility, and also offer the best justification and evaluative perspective on those rules and principles. But it is necessary to sound a note of caution: even theorists of the pure or mixed school might accept that criminal responsibility should not track moral culpability perfectly; such dislocation is arguably required in light of the criminal law's societal functions,⁸ the need for its reach to be defined with a reasonable degree of precision,⁹ and the constraints flowing from the practical reality of its administration.¹⁰ And the reference to societal functions reveals the possibility of an alternative approach to criminal responsibility that emphasises those functions and thereby considers it a virtue that there be some dislocation between criminal responsibility and the notion of moral culpability.¹¹ Such an approach seeks to re-conceptualise the criminal law's rules of responsibility in light of such societal functions. I shall examine such an approach below, when I explore the merits of Nicola Lacey's exploitation of character theory in the name of such a functional approach.

A further distinction requires articulation in order to understand those theories of criminal responsibility that are based on moral culpability. This is the distinction between the primary task of ascertaining moral responsibility for wrongdoing, and the secondary task, having established moral responsibility, of allocating blame amongst morally responsible wrongdoers. Ascertaining moral responsibility is the primary task because it addresses the fundamental attributes, possession of which makes a wrongdoer a legitimate candidate for blaming; once those attributes are more or less in place the secondary task of assessing the wrongdoer in light of his wrongdoing, and perhaps allocating blame in light of that assessment, may be entered into. This division of labour is mirrored in the context of exculpation by the distinction between exemptions and excuses: when an exemption is claimed the wrongdoer lacks one or more of the attributes that makes him a legitimate candidate for blame: it is therefore a *denial* of moral responsibility. On the other hand, when an excuse is claimed, the wrongdoer concedes his moral responsibility, but says that blame should be reduced, perhaps to extinction,¹² for certain reasons.¹³ To a

⁸ Such functions include incapacitation of the dangerous and general deterrence. Arguably these all flow from an umbrella duty of the state to maintain a safe social environment. Thus, for example, pure and mixed theorists might accept that the criminal law should not reflect moral culpability to such a degree that this seriously inhibits its capacity to deter.

⁹ Elements of offences that play a role in ensuring the moral culpability of the offender may have a detrimental effect on the precision of the criminal law. The requirement for dishonesty in theft is arguably an example: for a discussion of the requirement see Andrew Ashworth *Principles of Criminal Law* 6th ed (Oxford: Oxford University Press, 2009) pp 375-379.

¹⁰ Such practicalities arguably explain and justify crimes of strict liability. Thus moral culpability should not always be a necessary condition of criminal liability. Other liberal values may also restrict the criminalisation of morally culpable behaviour as pointed out by Moore: see *Placing Blame*, above n 5, Chapter 18.

¹¹ See Lacey, above n 6, p 68: "We must remind ourselves once again that an ascription of responsibility for the purposes of criminal justice is not identical with an assertion of purely moral blameworthiness."

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

large extent, therefore, a theory of moral culpability is a theory of exemptions and excuses, not least because, by understanding exemptions and excuses, one gains valuable insight into the positive grounds of moral responsibility and the allocation of blame.¹⁴ It will emerge during the course of this article that from both a descriptive and evaluative viewpoint I favour a theory of criminal responsibility based on a particular version of capacity theory. Rather than summarising my reasons in advance, I will let them emerge during the course of the analysis. In the next section, I will examine and reject character theory as a theory of criminal responsibility.

CHARACTER THEORY

Character Theory and Moral Culpability

As already indicated, for both pure and mixed theorists, criminal responsibility does and should track moral culpability. Some theorists believe that a theory based on character offers the best descriptive and evaluative conception of that moral culpability. This section is concerned with the plausibility of that belief.

The character theory's approach to moral culpability, specifically its approach to moral responsibility, is defined by the idea that there must be a link between the action of the wrongdoer and his character.¹⁵ More specifically, an agent will only be morally responsible for wrongdoing if it is in character; that is to say it must emerge from, or be in some way caused by or related to, a character trait of the agent. It is only under those circumstances that the agent is morally accountable for his actions. George Vuoso explains:

What an actor does is relevant to a moral evaluation of him to the extent that it reflects on the sort of person he is... Whether an action merits praise or blame, or reward or punishment, will depend on how it reflects on the agent, or on something enduring in the agent (which, following tradition, we are calling his 'character').... If an action is caused

¹³ Antony Duff 'Theorizing Criminal Law: A 25th Anniversary Essay' (2005) *Oxford Journal of Legal Studies* 353 at 361. See also Victor Tadros *Criminal Responsibility* (Oxford: Oxford University Press, 2005) p 124, section entitled: 'The Distinction between Excuses and Exemptions'.

¹⁴ As pointed out by Moore, see above n 5, p 548.

¹⁵ The following theorists have advanced versions of character theory, from rule-utilitarian, consequentialist or deontological perspectives: Duff 'Choice, Character and Criminal Liability' (1993) 12 *Law and Philosophy* 345; George Fletcher *Rethinking Criminal Law* (Boston: Little, Brown, 1978) pp 799-802 (though Fletcher ultimately abandons character theory); Michael D Bayles 'Character, Purpose and Criminal Responsibility' (1982) 1 *Law and Philosophy* 5; Tadros above n 13 (though character is a feature of a complex theory outlined by Tadros that draws on other accounts of moral responsibility in order to complete his account of moral and criminal responsibility); Lacey above n 6; George Vuoso 'Background, Responsibility and Excuse' (1987) 96 *Yale Law Journal* 1661; Lloyd Fields 'Moral Beliefs and Blameworthiness' (1994) 69 *Philosophy* 397; Robert Cummins 'Culpability and Mental Disorder' (1980) 10 *Canadian Journal of Philosophy* 207; and Richard B Brandt *Ethical Theory* (Englewood Cliffs, NJ: Prentice Hall, 1959) pp 465-474, 'Blameworthiness and Obligation' in A I Melden and Roger Squires (eds) *Essays in Moral Philosophy* (Seattle: University of Washington Press, 1958) p 3, 'The Utilitarian Theory of Excuse' (1969) 78 *Philosophical Review* 337, and 'A Motivational Theory of Excuse in the Criminal Law' in J R Pennock and J W Chapman (eds) *Criminal Justice, Nomos* 27 (New York: New York University Press, 1985) p 165. As Moore points out, this theory can be traced to David Hume, see above n 5, p 562, n 32.

or determined by the agent's character, it is clear that it reflects on his character: It was his character or some aspect of it that helped bring the action about.¹⁶

Thus character theorists believe that it is only that part of the agent that endures over time that is blamed for wrongdoing; Vuoso states "Actions are transient or fleeting things (though their effects may not be). The agent endures however."¹⁷ In turn, his character constitutes the part of the agent that endures over time. The rationale for this approach is that, so character theorists claim, it is only when wrongdoing is in character that it truly belongs to the agent, and only then is he a potential target for blame.¹⁸

Character theory has, at least at first blush, considerable descriptive power in explaining how the rules and principles of criminal responsibility track moral culpability.¹⁹ Thus mens rea concepts such as intention, knowledge and belief are significant for blame because, arguably, they act as conduits to the source of blame, the agent's malevolent or defective character.²⁰ By way of contrast, where excusatory defences such as duress or loss of control/provocation are concerned, the choice to do wrong is not revelatory of malevolent or defective character, with the result that blaming that agent is inappropriate.²¹ Similar explanatory potential applies to defences of an exempting nature. Thus bodily movements during an epileptic seizure, and other phenomena grouped under the defence of automatism, give rise to exemptions because they sever the link between the agent's character and wrongdoing: in the absence of that link, allocating blame is meaningless. Where the defence of insanity and those forms of mental disorder caught by the diminished responsibility defence are concerned, character theorists argue that possession of character requires, by definition, possession of an intelligible conception of reality and value, something that insanity and mental disorder denies.²² Thus the radically disordered thought patterns of the insane mean that they are not in possession of those otherwise culpable character traits, once again rendering the allocation of blame meaningless. Similar arguments inform the character theorist's account of the infancy defence: since children of a certain age have not had the opportunity to develop

¹⁶ *Ibidem*, at 1674.

¹⁷ *Ibidem*. See also Tadros above n 13, p 47: "When we are punished, we are punished as agents who persist over time".

¹⁸ See Lacey above n 6, pp 71-72. For a criticism of the metaphysical and moral basis of this rationale, see Moore above n 5, pp 580-581.

¹⁹ Character theorists who have explained the rules and principles of criminal responsibility in the terms that follow include Fletcher and Bayles, see above n 15.

²⁰ It should be acknowledged that some versions of character theory treat character rather than wrongdoing as the true object of blame: Lacey above n 6 is an example, as are Bayles and Duff, see above n 15. This makes character theory not a theory of responsibility for wrongdoing but rather a theory of criminal liability *tout court*. Such an approach needs to explain why the criminal law requires wrongdoing. In this regard, Duff has made the conceptual claim that a character trait is not in existence until it is manifested in action; for a criticism of this character-based solution to the requirement for action, see Moore above 5 pp 586-588. I agree with Moore's criticisms. As pointed out by Jeremy Horder, character theory is best seen as a theory of responsibility for wrongdoing and not a theory of what it truly is that we are blamed for: see 'Criminal Culpability: The Possibility of a General Theory' (1993) 12 *Law and Philosophy* 193 at 206.

²¹ See for example Duff above n 15, at 363 and Horder above n 20, at 205.

²² Duff above n 15, at 370: "...mental disorder is not the kind of 'defective character trait' that merits condemnation and punishment. For character, as an object of moral or legal criticism, consists in the person's rational dispositions – of thought, feeling and motivation; 'rational' here means roughly that his dispositions reflect an intelligible conception of reality and value."

characters in the full sense of the word, they are, to a greater or lesser degree, exempt from moral responsibility. According to character theorists therefore denials of moral responsibility, and their criminal law formulations, invoke some reason for thinking that the wrongdoing does not emerge from character: either because the agent is possessed of character, but somehow the link between that character and his actions (or more accurately his bodily movements or failures to act) has been broken, or because, at the relevant time, there is no character at all from which the action can flow. On the other hand, once the link to character is present, the agent can be evaluated in light of what the wrongdoing says about his character, and excused if that evaluation does not reveal (criminally) culpable character.²³

There is no denying the plausibility of character theory in describing how the existing rules and principles of criminal responsibility track moral culpability, though I will argue that, ultimately, character theory does not possess this descriptive power. My immediate concern, however, is to show how character theory suffers from a fatal flaw as a theory of moral responsibility that means it is not a suitable evaluative theory of criminal responsibility.

Key to the notion of character is the idea of consistency over time; as Antony Duff explains, character traits “are lasting or stable”, so that “a purely momentary feeling or impulse is no part of the person’s character... A character-trait... involves a pattern of thought, feeling and action, extending through time: it is not the kind of attribute which can be suddenly acquired and as suddenly lost, or which has just one single manifestation.”²⁴ It is this requirement for diachronic stability that ultimately undermines character as a theory of moral, and derivatively criminal responsibility. This is because, when combined with the requirement that wrongdoing always be in character, it leads on certain occasions to the wrong result or, if that conclusion is to be avoided, reveals character theory to be empty of content. The following example based on the crime of theft illustrates how this is so.

Imagine a rational agent, X, presented with an opportunity at a supermarket to appropriate an expensive bottle of Cognac without paying. X is affluent and free of addiction. X chooses to take advantage of that opportunity. However, as a matter of sincere evaluative belief, X considers theft morally wrong. He may have experienced some conflict as he stole the Cognac, or he may have taken it quite happily, excited by his own daring and by his violation of his principles.²⁵ I shall also assume that X has had similar opportunities for theft in the past, and it has not occurred to him to steal. And

²³ In light of the wealth of pejorative judgments that can be made of a morally responsible wrongdoer, it is worth noting that judgments about character can accommodate much moral subtlety and variation of subject matter where wrongdoing is concerned, embracing not only defective motivation (selfish, greedy, cruel, parsimonious) but also emotions (hot head) and lack of self-control (weak, gullible). Thus character theory possesses much descriptive power.

²⁴ Above n 15, at 364-365. See also Moore above n 5, p 582 and pp 586-587, and L C Holborow ‘Blame, Praise and Credit’ (1971) 72 *Proceedings of the Aristotelian Society* 85 at 93.

²⁵ Gary Watson argues that this is not necessarily a case of weakness of will: see ‘Scepticism about Weakness of the Will’ (1977) 86 *The Philosophical Review* 316 at 327, n 13, reprinted in R Jay Wallace (ed) *Reason, Emotion and Will* (Aldershot: Ashgate Publishing Company, 1999) p 427, p 438, n 13.

subsequently to this act of theft, X never steals again, despite a fair number of opportunities.²⁶ For X, the theft is an idiosyncratic one-off.

If it is accepted that X's theft is out of character, character theory's basic requirement of moral responsibility, that wrongdoing be in character, means that X should not be criminally responsible for his theft. This is intuitively the wrong result and is the reason character theory has been widely condemned as a theory of moral and criminal responsibility.²⁷ This condemnation points out, correctly, that the mere fact that wrongdoing is out of character because it is based on a fleeting motivation is insufficient to deny moral responsibility. In turn, if the criminal law were to acquit on such occasions it would rapidly fall into disrepute.

A way out of the above criticism is to treat chosen wrongdoing based on fleeting motivation as somehow being in character. This is difficult to argue because it is highly debatable whether character can be established, either constitutively or symptomatically, by a single instance of wrongdoing based on fleeting, idiosyncratic motivation. As we have seen above, character traits "are lasting or stable", so that "a purely momentary feeling or impulse is no part of the person's character...."²⁸ It would seem that the only way to guarantee that one-off wrongdoing such as X's is conclusive of character is to make the notion of character transient, a solution which contradicts its essential nature.²⁹ Nevertheless, Tadros and Duff have advanced arguments that supposedly accommodate responsibility for fleeting or idiosyncratic motivations within character theory. Tadros accepts that actions based on fleeting motivations are, in some sense, out of character, but argues that this is not sufficient to exculpate the wrongdoer under his refined version of character theory, which focuses on what Tadros calls the agent's 'persistent self'. Tadros explains:

That her action was motivated by an inclination that she does not commonly have is not sufficient to undermine the claim that her action was reflective of her persistent self. For her action was reflective of something central about her: not so much the persistence of her motivation, but rather her lack of desire or ability to control that motivation when it arises.³⁰

²⁶ Other examples of out of character wrongdoing that are nevertheless blameworthy are given by Moore, above n 5, pp 582-584.

²⁷ Examples of such condemnations include Moore, *ibidem*, pp 580-581; Horder 'Criminal Law: Between Determinism, Liberalism and Criminal Justice' (1996) 49 *Current Legal Problems* 159 at 171, and 'Criminal Culpability: The Possibility of a General Theory' above n 20, at 207; Alan Brudner 'A Theory of Necessity' (1987) 7 *Oxford Journal of Legal Studies* 339 at 345-7; Joshua Dressler 'Reflections on Excusing Wrongdoers' (1988) 19 *Rutgers Law Journal* 671 at 697-698; Wallace *Responsibility and The Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1996) p 112; Benjamin B Sendor 'The Relevance of Conduct and Character to Guilt and Punishment' (1996) 10 *Notre Dame Journal of Law, Ethics and Public Policy* 99 at 112-113; Holborow above n 24, at 94-95; Herbert Fingarette 'Rethinking Criminal Law Excuses' (1980) 89 *Yale Law Journal* 1002 at 1011; Carl Elliott *The Rules of Insanity: Moral Responsibility and the Mentally Ill Offender* (New York: State University of New York Press, 1996) pp 63-67; and Michele Moody-Adams 'On the Old Saw That Character is Destiny' in Amelie O Rorty and Owen Flanagan (eds) *Identity, Character, and Morality: Essays in Moral Psychology* (Cambridge, Mass: The MIT Press, 1990) p 111, p 119.

²⁸ Duff above n 15, at 364-365. See also Moore, above n 5, p 582 and pp 586-587, and Holborow, *ibidem*, at 93.

²⁹ A point made by Moore, *ibidem*, p 582.

³⁰ Above n 13, pp 52-53.

Duff makes a similar point, when examining the hypothetical case of a trusted employee who steals in his first and only act of criminal dishonesty:

It would be absurd to argue that we should not be morally or criminally liable for our weak-willed actions; and such actions surely reflect relevant character-traits. They show something about the nature and depth of our commitment to the values they flout; they show that we lack the moral strength to resist temptations that a more virtuous person would resist – or would not even be tempted by.³¹

However, there is sleight of hand in this argument, in the sense that both Tadros and Duff assume without justification that the failure to control the one-off motivation is indicative of a general tendency. But that general tendency may not in fact be present: it may equally be a one-off failure to control the impulse; there may not in fact be an *ongoing*, as Tadros puts it, “lack of desire or ability to control the motivation when it arises”, and if not ongoing, it is hard to see how it can be a feature of the agent’s persistent self. Nor, if the failure is a one-off, does it necessarily show, as Duff argues, “the nature and depth” of the agent’s commitment to the values he is flouting with his wrongdoing. One sees in responses such as this, as Michael Moore has expressed it, “a kind of fixed determination to find bad character expressed in any freely chosen bad action, no matter what.”³²

The manipulation of the concept of character to accommodate moral responsibility for fleeting motivation, and thereby extend moral responsibility under character theory to all circumstances where rational agents ‘freely’ choose their wrongdoing, reveals that the moral responsibility of such agents must reside elsewhere than in character. I shall argue that it resides in the agent’s possession of certain capacities, not least the capacity for moral understanding and, specifically where fleeting motivations are concerned, a capacity for self-control.³³ The upshot is that character theory is doing no work: it is an unnecessary and potentially misleading distraction where moral and criminal responsibility are concerned.

The above flaws of character theory are its flaws as a theory of moral culpability in the service of a theory of criminal responsibility that sees moral culpability as key to criminal responsibility. Such flaws would perhaps become irrelevant if the question of moral culpability was marginalised by an alternative, functional conception of criminal responsibility that emphasised the achievement of certain practical goals, and saw the notion of character as central to the achievement of those practical goals. It is to the plausibility of such an approach that I now turn.

Character Theory: A Functional Approach to Criminal Responsibility

Lacey believes that the criminal law should be exclusively interested in character and her reasons for so doing are derived from her justification of punishment and her rejection of

³¹ Above n 15, at 376

³² Above n 5, p 584. The following is perhaps an example of what Moore has in mind: “Furthermore...even actions which are out of character ordinarily reflect on the character of the agent” see Tadros, above n 13, p 296.

³³ See n 105 below and text.

rival justifications.³⁴ She rejects a version of the pure theory that justifies punishment upon the basis of the moral culpability involved in failing to exercise a capacity to do otherwise due to scepticism about the ability to formulate a meaningful theory of free choice and unexercised capacity.³⁵ She rejects the mixed theory because of the opposition it creates between punishment's utilitarian general justifying aim and its desert-based principle of distribution. Instead, she favours an exclusively functional theory that harmonises the general justifying aim of punishment with its individual distribution. She believes a theory focusing on character is that theory, because the general aim of punishment and its individual distribution are motivated by the same end, the targeting of those whose criminal actions are born of a settled disposition of "hostility or indifference towards, or rejection of, either that particular norm or the standards of the criminal law in general."³⁶

Lacey justifies the criminal law's exclusive focus on this character trait for a complex mix of functional reasons related to the proper goals of punishment as she sees them. Included in such goals are respect for autonomy, the promotion of social welfare, and also a measure of general deterrence and public protection. According to this approach, *mens rea*, and the various excuses and exemptions that populate the criminal law, filter out those possessed of the relevant undesirable character traits from those who are not.³⁷ At this point in the argument Lacey states:

This constitutes an internal relation between function and limitation: given the ultimate ends and functions of the criminal law, to serve the values of welfare and autonomy by specific means, the limitation of punishment to those who are dispositionally responsible for their offences is entailed both by the lack of need to punish others, whose offences are in some sense fortuitous, and by the different reactions we have to someone who does wrong as an aberration, and one who does so in a considered, characteristic way. In a community, it seems overwhelmingly likely that social cohesion, reaffirmation of common values, and denunciation would be best served in response to those whose actions exhibit a considered, settled rejection of community values or some aspect of them. Against other offenders, such a community can afford to (and therefore should, in accordance with the principle of urgency) adopt an attitude of toleration.³⁸

There are serious descriptive and evaluative problems with Lacey's approach. Descriptively, despite Lacey's claims, the criminal law does not always distil those possessed of the relevant undesirable character trait from those not so possessed. This is because, in light of the possibility of out of character wrongdoing outlined above, possession of *mens rea*, and the absence of defence, does not guarantee that criminal wrongdoing is an expression of an undesirable character trait, let alone the specific

³⁴ See generally *State Punishment*, above n 6. A similar approach can also be seen in Bayles, above n 15.

³⁵ *Ibidem*, pp 62-65.

³⁶ *Ibidem*, p 76. Bayles also speaks in terms of teasing out a particular undesirable character trait, in the form of a trait "likely to produce criminal harm": see above n 15, at 13. It is worth noting that Lacey believes that her harmonised approach retains a culpability element ("backward-looking element") that respects individual autonomy: see p 72.

³⁷ *Ibidem*, p 76 and p 190.

³⁸ *Ibidem*, p 190.

negative trait that Lacey is interested in. The result is that the criminal law is a very blunt instrument indeed if its justifying rationale is to seek out those possessed of Lacey's undesirable character trait. Lacey might respond in two ways to such an argument: she could insist that intentional wrongdoing and the absence of defence guarantees the presence of the relevant character trait: this smacks of dogmatism and, as argued above, empties character theory of content; alternatively, she might accept the argument and advocate further reform of the criminal law to fine-tune its ability to pick out the relevant character defect.

From an evaluative viewpoint, this remains undesirable: for example, as Moore's analysis of Richard Herrin's murder of Bonnie Garland reveals, having established that a person has killed intentionally for some morally repugnant purpose, a further exploration of whether this was in fact in character, and an acquittal if concluded that it was not, seems equally morally repugnant.³⁹ Lacey can nevertheless stick to her guns here, and say that, despite such morally undesirable consequences on certain occasions, such a system of criminal law is the product of the best justification of punishment. If so, attention turns to the practicalities of such a system.

Surprisingly, character theorists have more or less ignored the question of their theory's practical implications. Thus Bayles, one such theorist, argues that the criminal law presumes that the presence of mens rea, and the absence of defence, filters out those possessed of the relevant undesirable trait(s) without, as Duff has pointed out, saying how that presumption is justified.⁴⁰ Similarly, Fletcher has argued in an exploration of the merits of character theory that this failure to investigate further into character is justified in the name of privacy.⁴¹ This approach leaves character theory vulnerable to the descriptive failures outlined above. Lacey has addressed the question more directly, suggesting that the courts will often need "to look more broadly at the defendant's attitudes as manifested in other relevant areas of behaviour" and also "to broaden the focus of its time-frame backwards to earlier stages (and even forwards, in cases of apparently genuine remorse)" to establish "the inference from action to disposition"⁴², but does not explore how this is to be done. The issue here is that there are significant conceptual and practical problems with this broader inquiry.⁴³

From a conceptual viewpoint, assuming satisfactory mechanisms for gathering the potentially relevant evidence, the task would often prove hopelessly vague, as the court explored that evidence in order to decide whether the accused possesses Lacey's criminal character trait. Such vagueness at the point of conviction would arguably lead to the inconsistent treatment of defendants and risk the manifestation of prejudice. Furthermore, it might very well have an adverse effect on the criminal law's capacity to deter, at least for those who saw an opportunity for a good character defence. There are also numerous practical difficulties with such a process. First, it would be resource intensive. Secondly,

³⁹ An analysis in direct response to the implications of Lacey's argument: see *Placing Blame* above n 5, pp 578-580.

⁴⁰ Above n 15, at 10. For Duff's comment see above n 15, at 374.

⁴¹ Above n 15, pp 800-801.

⁴² All the quotes in this sentence come from *State Punishment*, above n 6, p 75.

⁴³ Duff above n 15, at 368: "...the criminal process is radically ill-suited to serve as an inquiry into the subtler depths of a defendant's moral character; nor is such an inquiry the law's proper business." I do not believe that this problem of vagueness is avoided by narrowing the sought-after character trait to a disposition to break the law.

in light of the current limitations on admission of bad character by the laws of evidence,⁴⁴ trials would have to be radically reformed so as to proceed in two stages: a first stage dedicated to ascertaining whether the accused committed the criminal act in question free of (prejudicial) reference to bad character,⁴⁵ and a second stage dedicated to establishing criminal responsibility once such ‘guilt’ had been established, where this limitation would be lifted in order to decide whether the crime was the product of the targeted criminal character trait. Thirdly, who should perform this second stage where trial by jury is concerned? If left to the common sense of the jury, perhaps after listening to expert evidence, inconsistent results could flow in light of the conceptual vagueness outlined above. In an attempt to avoid this danger, the task could be given to judges and/or experts; the problem with this solution is that the democratic security offered by the jury would be lost.⁴⁶

This is not to say that character should not have a role to play after conviction, as a factor in sentencing. The aim of sentencing is to impose the appropriate sentence, a matter of degree that requires responding to a range of factors, not least offence gravity. Given this opportunity afforded by sentencing for greater subtlety in the assessment and treatment of the offender, it is not inarguable, in light of the sentencing ideals of incapacitation, rehabilitation, and even retribution, to allow character to influence mode and length of punishment.⁴⁷ However, such a range of factors cannot be accommodated at the point of conviction since conviction is a threshold question designed to decide whether punishment is deserved at all, and therefore has an all-or-nothing quality ill-suited to the vagaries of the notion of character.⁴⁸

It is submitted that, in light of these serious practical difficulties, and also the potential for undesirable moral outcomes, Lacey’s functional vision of the relevance of character to criminal responsibility should be rejected. However, before turning my attention to what I believe is the more promising theory of criminal liability, capacity theory, I wish to explore one final argument that seeks to make character relevant to criminal responsibility. Unlike the theories explored so far, this approach does not offer character as a universal theory of criminal responsibility but locates its relevance in a more limited way.

Alienation Theory: A Role for Character?

Despite character theory failing as a universal theory of criminal responsibility, perhaps certain exculpatory claims need the notion of character nevertheless. This possibility of a more limited role for character is linked to scepticism about the possibility of creating

⁴⁴ I confine my comments here to English law.

⁴⁵ Admittedly the Criminal Justice Act 2003 has expanded the ability to refer to bad character in order to establish whether the accused committed the criminal act: see ss 98, 101 and 112. But Lacey’s broader inquiry into character would almost certainly embrace evidence that remains inadmissible under this new legislation, necessitating the two-stage trial process articulated in the text above or, alternatively, a radically new approach to the admissibility of bad character evidence.

⁴⁶ Where trial in the Magistrates’ courts is concerned, the whole task would have to be performed by magistrates, a complex factual and conceptual task, and somewhat strange where minor offences are concerned.

⁴⁷ See Ashworth ‘Taking the Consequences’ in S Shute, J Gardner and Horder (eds) *Action and Value in the Criminal Law* (Oxford: Clarendon Press, 1993), p 113.

⁴⁸ See generally Gardner ‘Crime: In Proportion and in Perspective’ in *Offences and Defences: Selected Essays in the Philosophy of the Criminal Law* (Oxford: Oxford University Press, 2007).

universal theories of criminal responsibility. This alternative approach rejects the wisdom of uniting all exempting and excusatory claims under a single conceptual banner, but instead designs individual exempting and excusatory claims based on capacity, or character, or some other rationale, as appropriate.⁴⁹

Tadros supports this eclectic approach in the context of his vision of criminal law excuses,⁵⁰ the role of which he conceives as being "...to sweep up cases where the defendant has committed the relevant act, but where his action is an inappropriate target of criminal responsibility."⁵¹ Tadros believes that character theory offers one of these non-exclusive reasons to excuse, specifically the notion of *alienation from character*. What I will examine is the plausibility of Tadros's belief that his alienation theory provides the best account of the excusatory power of the English common law formulation of the defence of provocation, now abolished.⁵² In order to explain the rationale behind alienation theory, Tadros quotes G R Sullivan:

[T]here are occasions when criminal liability should be precluded by invoking some version of 'unity of the self' doctrine informed by criteria going beyond mere bodily continuity. Such a doctrine must hold that certain core values subscribed to by a particular agent are constituents of selfhood and that any sudden and fundamental change in those values for which the agent is not responsible makes for a discontinuity of the self.⁵³

I noted above that at the heart of character theory's approach to criminal responsibility is the notion that the agent is only criminally responsible for wrongdoing when it emerges from, or is causally related to, his character. Tadros subscribes to a version of this idea whereby an agent is criminally responsible for wrongdoing when the wrongdoing "reflects on the agent *qua* agent" and this is the case when there is a connection between the wrongdoing and the agent's character.⁵⁴ In turn, Tadros interprets Sullivan's claim that discontinuity of self should, when it occurs non-culpably, preclude criminal liability as an instance of this general claim that criminal responsibility is not appropriate when the link between character and wrongdoing is not in place.⁵⁵

⁴⁹ See generally Horder 'Criminal Culpability: The Possibility of a General Theory' above n 20.

⁵⁰ Above n 13, Chapter 11 'The Characters of Excuse'.

⁵¹ *Ibidem*, p 294. This is a concept of excuse that embraces any good reason for creating a criminal law defence when those reasons are not concerned with a fundamental incapacity for rational action.

⁵² Above n 13, pp 302-306. The common law defence of provocation was abolished by s 56 of the Coroners and Justice Act 2009. The statutory 'loss of control' defence that replaced it is defined by ss 54 and 55 of the Act. The common law defence was itself modified by s 3 of the Homicide Act 1957, which, in light of the abolition of the common law defence, has ceased to have effect. See below n 109 and n 111 for more discussion of the 'loss of control' defence.

⁵³ 'Making Excuses' in Andrew Simester and A T H Smith (eds) *Harm and Culpability* (Oxford: Clarendon Press, 1996) p 131, p 137.

⁵⁴ See p 47 and p 297, above n 13.

⁵⁵ *Ibidem*, p 297: "Where the agent has undergone a fundamental shift in character for which he is not responsible, the actions performed as a consequence of that shift do not reflect on the agent *qua* agent in the appropriate way."

Key to this approach is a definition of the self: according to Tadros, selfhood is constituted by what he terms ‘settled character’,⁵⁶ which he defines as being...

... made up, at least to a significant degree, by the reasons that one is commonly moved by. Hence, the brave are not commonly deterred by danger but rather are challenged by it; the selfish are commonly motivated by their own interests rather than the interests of others and so on. We might call this one’s motivational set up.⁵⁷

Before proceeding to analyse how Tadros applies his alienation theory to the common law version of provocation, a preliminary criticism questions why the agent’s identity or selfhood should be exclusively constituted by those motivations that commonly move him. A weakness of Tadros’s approach, flowing from its dependence on a particular conception of the agent, is that its ultimate success depends on the resolution of the complex debate about which motivations are correctly identified with the agent, and which are not.⁵⁸ If Tadros’s definition of the agent in terms of settled character is too narrow, the agent is *not* alienated from his ‘self’ when he acts for reasons that do not commonly move him, and his theory collapses from within. But we can give Tadros the benefit of the doubt here, and identify the agent with his settled character as Tadros conceives it. The question therefore is whether his conception of alienation from character gives the best account of the common law defence of provocation.

Tadros focuses on the common law defence’s requirement for a loss of self-control. He interprets this as involving the agent acting for reasons that would not ordinarily move him, thereby bringing the defence within the ambit of alienation theory. That this happens is due to the effect of anger; Tadros goes on to say:

If this explanation of the ‘loss of self-control’ element of the defence of provocation is correct, we can see that it has its basis in the character theory of excuses that I have been defending here. The reason for introducing this element is that the excuse will only be available insofar as the agent was put into a state in which her ordinary reasons for action did not guide her action in the way that they normally would. Her character whilst she is in a state of extreme anger is not like her character whilst calm. Actions done in that state do not reflect as badly upon her settled character as if they had been done whilst calm. But...this is only so if the settled character of the agent is peaceful. If it is not, then she cannot show the difference between her settled character and her outraged character that provides the basis of the defence.⁵⁹

This is undeniably plausible, but there is more to provocation as an excuse than merely losing self-control, and thereby acting for reasons that do not commonly move one. There is an evaluative element, captured in the common law defence’s further requirement that

⁵⁶ *Ibidem*, p 298. There is a tension between Tadros’s identification of the agent with his settled character and Sullivan’s identification of the agent with his core values, as some agents, for example the chronically weak willed, are commonly motivated by reasons that are in conflict with their core values. For Tadros, such agents are nevertheless acting in accordance with their settled character.

⁵⁷ *Ibidem*.

⁵⁸ See Moore *Placing Blame* above n 5, pp 580-581, where this criticism is leveled at Lacey’s view that actions only truly belong to an agent when they are in-character.

⁵⁹ Above n 13, pp 303-304.

the loss of self-control be reasonable.⁶⁰ It is when this additional excusatory factor is explored that Tadros's explanation of provocation as an excuse cannot be sustained. Tadros recognises, correctly in my view, that the agent's reasons for acting, however grave the provocation, have no capacity, *qua* reasons for acting, to reduce blame for killing. Tadros explains:

The difficulty is that in cases of provocation *there is nothing at all* to be said in favour of the action that the defendant has performed [author's emphasis]. There is nothing to justify killing under provocation. That one is provoked is not even an insufficient reason to kill. It is no reason at all.⁶¹

Thus the fact that the wrongdoing is a response to a grave provocation, in and of itself, is insufficient to excuse, even partially. Something else requires articulation if we are to understand why we excuse in such cases. Tadros sees that something else in the fact that the grave provocation means the agent is alienated from his settled character in a justified way, thereby accounting for the evaluative component of the defence:

But the justification ought to go not to the action itself but rather to the loss of self-control. ...[T]he accused has a justification for getting into that state. This rests on the now familiar argument that it is sometimes not only *natural* to become angry, it is also right to become angry [author's emphasis]. But once one is extremely angry, one's actions may no longer reflect one's settled motivational set quite as closely as they do when one is calm. This, in my view, provides the best conceptual foundation of the defence of provocation.⁶²

There are two major problems with Tadros's argument. First, Tadros seems to believe that being angry for good reasons is all that is required for provocation to function as an excuse. His argument thereby dislocates the agent's reasons for being angry from the reasons for which the agent acts. He loses sight of the fact that, despite the failings of the agent's reasons for acting, the excuse nevertheless depends on the agent *reacting* to the provocation. Otherwise, any wrongdoing by the agent once he is angry for good reasons would be excused. For example, the agent may not only kill the provoker but also the provoker's friend standing nearby, out of spite. If the agent was justifiably angry and thereby alienated from his settled character at that moment, then, according to Tadros, he should be excused, and clearly he should not be.

Second, the inherent logic of character theory does not justify this requirement for a *justified* alienation from settled character. Character theory seeks to deny the link to character, which is an empirical question. Thus the mere fact of alienation from settled

⁶⁰ The new statutory defence of 'loss of control' created and defined by the Coroners and Justice Act 2009, see above n 52, arguably incorporates this requirement. This is because its concept of a qualifying trigger necessary for the defence to be available, set out in s 55(3) and (4), is constituted either by, on the one hand, the defendant's "fear of serious violence" against himself or another identified person or, on the other, "a thing or things done or said (or both)" of an "extremely grave character" which cause the defendant to have "a justifiable sense of being seriously wronged". Finally, under s 55(5), a combination of the matters mentioned in s 55(3) and (4) will also constitute a qualifying trigger.

⁶¹ Above n 13, p 304.

⁶² Above n 13, pp 304-305.

character should suffice to preclude criminal liability under character theory, without any analysis as to why. This is no doubt why Tadros endeavours to explain the requirement for justified anger purely in terms of the empirical question of attribution to character:

D is in a state, *x* (say anger [...]), which distorts his motivations: the normative reasons that would ordinarily motivate his actions do not motivate his actions in the normal way. Consequently, actions in *x* do not reflect his settled character as closely as actions whilst not in *x*. Ordinarily, however, D has no excuse. Reprehensible actions in *x* may not reflect as closely his settled character as actions when he is settled, but *that* he is in *x* is attributable to him [author's emphasis]. This is why we tend not to excuse those who respond angrily to trivial provocation... However, if D is justified in being in *x*... then he is entitled to an excuse as neither his actions in *x* nor being *x* reflect sufficiently badly upon his settled character.⁶³

But this is suspect for two reasons. First, just because an agent reacts badly to trivial provocation does not mean his settled character is expressed in so doing: it may be an out-of-character one-off. Secondly, actions in justified anger will more often than not reflect on settled character as much as those resulting from unjustified anger, because the agent's most important values will explain why he is angry.⁶⁴ What the talk of justified anger reveals is that attribution to settled character is irrelevant; what matters is the quality of the agent's action (and anger) in light of an objective standard. As pointed out by Horder, Tadros himself slides into such a claim when he states that actions in instances of justified anger do not reflect sufficiently *badly* on settled character, rather than not reflecting at all.⁶⁵ The question of attribution to settled character is therefore beside the point in distilling meritorious cases from the unmeritorious. What is required is an approach that accommodates both the effect of anger on motivation and the evaluative elements necessary to explain provocation as excuse: as I shall argue below, this is achieved by conceiving of provocation in terms of an evaluation of the exercise of self-control.

I now turn my attention to what I believe is a far more promising theory of criminal responsibility, the capacity theory.

CAPACITY THEORY

Criminal law theorists who reject character theory as the means of explaining how the rules and principles of criminal responsibility track moral culpability have sought to explain and evaluate those rules and principles in terms of the notion of capacity. Capacity theory locates an agent's moral responsibility in his possession of certain capacities and allocates blame in light of the conditions under and the reasons for which those capacities are exercised. However, the exact applicability and meaning of this general statement depends on the version of capacity theory under analysis.

⁶³ *Ibidem*, pp 305-306.

⁶⁴ See Horder *Excusing Crime* (Oxford: Oxford University Press, 2007) pp 123-124: "There is no analogy with the effects of involuntary intoxication, which exercise their influence on D in a causally passive manner, in that respect completely by-passing the kind of actively interpretive (character-based) frame of reference through which words or actions come to be understood by a D as 'provocative'".

⁶⁵ *Ibidem*, p 124.

All versions of capacity theory offer some conception of the capacity for moral understanding required to constitute the wrongdoer as morally responsible. The different versions of capacity theory conceive of the required moral capacity at various points on a spectrum of competence. I will begin my analysis of capacity theory by setting out the account of that competence that I favour, showing how that account underpins moral responsibility and enables the rules and principles of the criminal responsibility to track moral culpability. I will then proceed to examine and reject certain versions of the theory. In so doing, all the features of the capacity theory that I favour will emerge.

The Relevance of Moral Capacity

The following definition by Duff of the required moral capacity is exemplary:

[Moral capacity] involves more than the ability to apply learned formulae which provide descriptive criteria for the identification of actions to which moral labels can then be attached. ... A person who is to exhibit moral understanding must also be able to explain and criticize these moral rules - which involves more than showing how they do or do not derive from other formulae; he must be able to show how these rules may or may not be *extended* to cover new cases, which do not fall exactly under any specified set of descriptive criteria [author's emphasis]; to discuss rationally the resolution, or the impossibility of resolving, cases of conflict. Following moral rules, the ability 'to go on in the same way', requires more than the intellectual capacity to acquire and apply fixed formula: it requires a *creative* capacity to understand the significance of the value in question and to discuss, extend and criticize its application [author's emphasis].⁶⁶

Why is such a capacity required? The answer lies in focusing on the fact that blame is a pejorative moral judgment, in the form of the conclusion that the agent's wrongdoing casts him in a pejorative moral light. This judgment is allocated after a process of moral evaluation that assesses the moral quality of the wrongdoer's reasons for acting in light of his failure to be decisively motivated not to do wrong by the reasons against his behaviour. When the agent's reasons for acting prove inadequate in light of the reasons against, the agent is blameworthy. He has made a moral error in failing to accord the correct moral weight to those reasons, and that moral error grounds his blame.⁶⁷ Inherent to this process of appraisal is the fact that the agent's choices are *morally meaningful*. What this means is that the agent must understand the moral significance of his behaviour as this grounds the logic of the appraisal described above.

Examining excuses, the conditions under which morally responsible wrongdoers can negate blame for wrongdoing, reinforces the above account of the significance of moral capacity. Much (though not always all) of an excuse's exculpatory power lies in the wrongdoer being able to give a morally intelligible account of his action: to give such an account is not for the wrongdoer to claim that his action is justified (his reasons for doing as he did are not adequate for that purpose), but rather to claim that his reasons for doing

⁶⁶ 'Psychopathy and Moral Understanding' (1977) 14 *American Philosophical Quarterly* 189 at 195. I believe this capacity involves a complex mixture of cognitive and emotional capacities.

⁶⁷ Larry Alexander captures this moral error in the notion of "insufficient concern for the legally protected interests of others": see his *Crime and Culpability*, written with Kimberley Kessler Ferzan and Stephen Morse (Cambridge: Cambridge University Press, 2009), Chapter II 'The Essence of Culpability'.

as he did exhibit certain rational features such that his wrongdoing does not reflect so badly on him.⁶⁸ Thus to evaluate a wrongdoer's reasons for acting in order to decide whether to excuse him necessarily entails a wrongdoer who, in offering those reasons, is endeavouring to make a certain kind of sense of his actions.⁶⁹ Agents who are incapable of making sense of their wrongdoing in this way cannot advance excuses: this is why offering an excuse presupposes a moral capacity.⁷⁰

In light of the above, the absence of moral capacity at the moment of wrongdoing, in the form of irrationality (insanity) or non-rationality (infancy), constitutes an exemption. This is because it makes no sense to evaluate morally the choices of agents whose choices are morally meaningless. Capacity theory's explanatory power where such exemptions are concerned is deeper and more illuminating than character theory, which merely denies the presence of character on such occasions. The exemptions of insanity and infancy, in morality and law, are best understood as involving a lack of capacity rather than a denial of character because the significance of lack of character as a denial of moral responsibility lies in the absence of the capacity for moral understanding.

In what sense do the above insights enable the rules and principles of criminal responsibility to track moral culpability? Where *mens rea* is concerned, the presence of intention, recklessness, belief etc indicates the wrongdoer's choice to commit the wrong: when this fact is married with a capacity for moral understanding, the choice, all other factors being equal, is constitutive of the moral failure described above,⁷¹ and thus the wrongdoer's blameworthiness. We have seen that to benefit from an excuse, the agent must account for his actions in such a way as to negate, in whole or in part, the moral failure potentially present in his chosen wrongdoing. This is done to a large extent by evaluating his intelligible motivation. Thus it is interesting to note that integral to formulations of the defences of provocation and duress are the agent's intelligible reasons for feeling and/or acting as he did, filtered by the requirement that those reasons exhibit certain positive moral features (a grave provocation, the avoidance of death or serious physical injury injury).⁷² Where exemptions are concerned, the defence of infancy is also best explained in terms of the criminal law's acknowledgment that children lack a fully developed moral capacity, and that their consequent lack of moral responsibility should be reflected by the criminal law in the form of a full defence.⁷³

The most obvious candidate for explanation in terms of a capacity for moral understanding is the insanity defence, which most often takes the form of the M'Naghten

⁶⁸ Gardner 'The Mark of Responsibility' (2003) *Oxford Journal of Legal Studies* 156, at 156-159. See also Horder *Excusing Crime* above n 64, pp 45-52.

⁶⁹ Tadros, above n 13, p 124 "An agent who is exempt from responsibility is not an appropriate agent from whom we can demand an explanation." This is true, and the reason lies in his very inability to offer that explanation.

⁷⁰ See generally Gardner 'The Mark of Responsibility' above n 68.

⁷¹ See above n 67 and corresponding text.

⁷² See, for example, the qualifying triggers for the new partial 'loss of control' defence to murder set out in s 55 of the Coroners and Justice Act 2009: see above n 60. The avoidance of death or serious physical injury plays a key role in the defence of duress under both English and American law; for English law see n 90; for the US see the Model Penal Code, adopted purely or adapted by many US states, which refers to the use of, or the threat to use, 'unlawful force' against the accused or the person of another; see section 2.09.

⁷³ The law must draw the line somewhere: in England the age is 10 (s 50 of the Children and Young Persons Act 1933, as amended); the age varies in the US depending on the State concerned.

rules.⁷⁴ However, in light of the current interpretation of these rules, it is best if I address the implications of capacity theory for them below, when I explore the ‘choice’ version of capacity theory.

The above demonstrates the descriptive and evaluative power of a theory that focuses on moral capacity in explaining how the criminal law tracks moral culpability. But not all capacity theorists conceptualise the moral competence necessary to ground moral responsibility in the same way. It is to one of these conceptions that I now turn, which I have termed the good moral motivation theory.

Rejecting the Good Moral Motivation Theory

According to good moral motivation theorists,⁷⁵ in order to be morally responsible the agent must not only possess the capacity for moral understanding as outlined above, he must also, as Susan Wolf expresses it, possess the “ability to recognise good values as opposed to bad ones and to act in a way that expresses appreciation of this recognition. [He must possess] the freedom and power to do the right thing for the right reasons.”⁷⁶ Peter Arenella explains further:

What *sustains* our [blaming] attitudes is not simply the actor’s breach of the moral norm, but our interpretation of the actor’s lack of concern or respect for the norm breached [author’s emphasis]. But, our interpretation of the actor’s attitudes towards these norms presupposes that we are dealing with a human being who has the capacity to understand their significance and the ability to respond appropriately to their demands. When an individual who engages in morally objectionable behaviour lacks this understanding or responsiveness, it becomes more difficult to interpret his acts as demonstrating a *culpable* lack of concern for the norm [author’s emphasis].⁷⁷

Good moral motivation theorists therefore believe that the ascription of blame when the agent’s motivation is morally defective only makes sense when the agent did or could know better, and could have acted differently in light of that actual or potential knowledge. Thus, according to good moral motivation theorists, the ascription of blame requires a failure of self-control of some description in order to be meaningful. The descriptive failures of the good moral motivation theory where the criminal law is concerned are clear: there are no defences based on lack of capacity for good moral

⁷⁴ Many US states use the M’Naghten rules. It is the legal test for insanity under English law (10 Cl & Fin at 210): “it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong.”

⁷⁵ The following authors have outlined versions of the good moral motivation theory: Susan Wolf *Freedom Within Reason* (Oxford: Oxford University Press, 1990); Wallace *Responsibility and The Moral Sentiments* above n 27; Peter Arenella ‘Character, Choice and Moral Agency: The Relevance of Character to Our Moral Culpability’ (1990) 7 *Social Philosophy And Policy* 59, reprinted in Ellen F Paul et al (eds) *Crime, Culpability and Remedy* (Oxford: Blackwell, 1990) p 59 and ‘Convicting the Morally Blameless: Reassessing the Relationship Between Legal and Moral Accountability’ (1992) 39 *University of California Los Angeles Law Review* 1551; Jeanette Kennett *Agency and Responsibility: A Common-sense Moral Psychology* (Oxford: Clarendon Press, 2001); and Philip Pettit and Michael Smith ‘Freedom in Belief and Desire’ (1996) 9 *Journal of Philosophy* 429.

⁷⁶ *Freedom Within Reason*, *ibidem*, p 77.

⁷⁷ Arenella ‘Convicting the Morally Blameless...’ above n 75, at 1539.

motivation. Admittedly poor social background, when resulting in an inability to react to the persuasive force of certain reasons against wrongdoing, may serve as mitigation in the sentencing process, but there are no defences effective at the point of conviction based on this rationale. But those who have defended the good moral motivation theory have done so more in the spirit of critical evaluation than an attempt to describe the criminal law's actual practices. The question, therefore, is whether the criminal law would be best reformed in light of their insights.

I think not. From a moral viewpoint, the claim that moral responsibility requires a failure of self-control is highly debateable: blame is also meaningful when the agent's wrongdoing emerges out of a personal scheme of values in opposition to good values, in the absence of any capacity for revision. In other words, all other factors being equal, the fact that an agent is evil as opposed to weak-willed should have no bearing on his moral responsibility.⁷⁸ The failures of the good moral motivation theory as an evaluative theory become even more acute when contemplating the institutional and societal role of the criminal law: creating defences designed to prevent convictions due to the agent's inability to do the right thing for the right reasons, especially if those defences were carved out of socio-economic factors as opposed to medical grounds, would arguably bring the criminal law into disrepute and be potentially very harmful to its communicative and deterrent effects. It is for these reasons that the good moral motivation theory should be rejected as a theory of criminal responsibility.

I will now focus on what is perhaps the best-known capacity theory: choice theory. It is important to note two key features of choice theory. First, it is characterised by a very thin account of the normative competence necessary to ground moral and criminal responsibility. Secondly, it seeks not only to account for the moral competence necessary to ground moral responsibility but to offer a global theory of criminal responsibility grounded in moral culpability.

Rejecting Choice Theory

In determining criminal responsibility, choice theory focuses on the agent's capacity for choice and the opportunities he has for exercising that capacity. If the agent's choosing capacities are intact, and his opportunities for exercising those capacities are not unfairly diminished, such an agent is legitimately held criminally responsible if he chooses to violate the criminal law. In turn, criminal law defences, whether exempting or excusatory in nature, are carved out of the notion that the agent did not have sufficient capacity or opportunity to choose to do otherwise.⁷⁹

Key to choice theory is its conception of the process of choice necessary to ground moral responsibility. According to choice theorists, it is present when the agent selects his

⁷⁸ For an expression of the idea that incontinence is not as bad as evil, see Donald Davidson *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), pp 24-25. See also Holly Smith 'Varieties of Moral Worth and Moral Credit' (1991) 101 *Ethics* 279 at 296, where she argues that under what she terms the 'moral unfitnes' test of blame, a person whose commitment to morality is stronger appears *less* blameworthy than the agent with the weaker commitment.

⁷⁹ A classic formulation of the capacity theory is to be found in the writings of Hart: see generally *Punishment and Responsibility*, above n 6, specifically p 152. For an analysis of Hart's justifications of his formulation of capacity theory, see Moore *Placing Blame* above n 5, pp 549-552.

wrongdoing after a process of undisturbed practical reasoning.⁸⁰ Robert Schopp, a choice theorist, explains: “The actor acts as a practical reasoner when he deliberates regarding possible action-plans in light of a comprehensive network of wants and beliefs, selects an action-plan intended to maximise his want satisfaction, and acts according to his action-plan in a manner such that his acts are caused by his wants and beliefs according to the usual process.”⁸¹ He continues: “An effective process of practical reasoning requires not only access to the relevant wants and beliefs, however, but also intact capacities of concept formation, comprehension, and reasoning for selecting the action-plan likely to promote the set of wants. In order to be effective, the action-plan selection process requires at least the capacities to recognise intuitive implication and inconsistency and to maintain a set of beliefs with a satisfactory degree of coherence.”⁸² Even so, as Arenella has pointed out, the required capacity is one exclusively constituted by the agent’s instrumental reasoning capacities, with a very low threshold of rationality.⁸³ So long as the agent could choose otherwise in light of some intelligible goal, which could be confined to self-interest alone, he has sufficient normative capacity for criminal responsibility, even if he lacks any comprehension of the moral significance of his wrongdoing. Such an approach means that, as Arenella expresses it, only the very crazy or very young will lack the required instrumental reasoning capacity, with the result that exemptions will be confined to such agents.⁸⁴

Choice theory’s vision of the normative competence required to ground criminal responsibility has descriptive power where the English definition of insanity is concerned. This is because the first limb of the M’Naghten rules is confined to those agents whose incapacity is cognitive as opposed to normative (a failure to understand the physical nature of the act), and the second limb has been interpreted to confine the relevant normative incompetence to the inability to appreciate the fact of criminal, as opposed to moral, prohibition.⁸⁵ Thus, for example, if the defendant believes his action is morally justified due to the effect of paranoid schizophrenia this will not avail him if he is

⁸⁰ It is worth noting that character theorists such as Vuoso and Kyron Huigens have interpreted Moore’s requirement for undisturbed practical reason as being significant because, so their argument goes, when practical reason is undisturbed, wrongdoing is the product of, or reflects on, the agent’s character. For Vuoso see above n 15 at n 36; for Huigens see ‘Virtue and Inculpation’ (1995) 108 *Harvard Law Review* 1423 at 1441-2. However, as pointed out in this article, this is not always true: practical reason can be undisturbed without the wrongdoing that emerges from it necessarily issuing from, or reflecting on, an agent’s character: see above n 28 and surrounding text.

⁸¹ *Automatism, Insanity, and the Psychology of Criminal Responsibility: A Philosophical Enquiry* (Cambridge: Cambridge University Press, 1991) p 191.

⁸² *Ibidem*, p 195

⁸³ Above n 75, at 65.

⁸⁴ *Ibidem*.

⁸⁵ See above n 74. For England and Wales see *R v Windle* [1952] 2 QB 826, confirmed in *R v Johnson (Dean)* [2007] EWCA Crim 1978; [2008] *Criminal Law Review* 132. American courts are divided on this particular question: see Sanford H Kadish, Stephen J Schulhofer and Carol Steiker *Criminal Law and its Processes; Cases and Materials* 8th edition (Aspen Publishers, 2007) pp 889-893 ‘The Meaning of Wrong’. Under English law awareness that the act is morally wrong will also result in the defence failing even if the accused is unaware of the fact of legal prohibition: see *Codère* (1916) 12 Cr App R 21 at 27. This latter position strikes me as almost certainly correct (after all ignorance of the law is no excuse) though much depends on the nature of the wrongdoer’s understanding of the moral wrongness of his act and why he is ignorant of legal prohibition.

aware of the fact of legal prohibition.⁸⁶ But this current interpretation of the second limb of the M’Naghten rules represents an important failure by the English criminal law to track moral culpability. As a result, it should be altered so that the second limb captures not only the inability to recognise the fact of criminal prohibition, but also the moral significance of wrongdoing.⁸⁷

Further understanding of choice theory is obtained by an examination of Moore’s account of excuses, as Moore is the leading exponent of choice theory.⁸⁸ Moore conceptualises excuses as either instances of interfered with/disturbed practical reason (defective capacity) or as occasions when opportunities to exercise choice are unfairly diminished (diminished opportunity).⁸⁹ This approach has immediate descriptive appeal where certain criminal law defences are concerned: excuses such as duress and provocation would appear to impact on either the agent’s capacity for choice, or his opportunities to exercise that capacity; exemptions such as automatism on the capacity to choose at all. However, this appeal is superficial: if one scratches the surface, it emerges that choice theory offers too thin an account of the capacities necessary both to explain and evaluate these defences. This is due to its attempts to dislocate the agent’s choosing capacity from the agent’s underlying values and understanding of normative questions generally. I will explain.

Moore realises, doubtless because of the frequent references to emotion in formulations of the provocation and duress defences,⁹⁰ that his defective capacity conception of excuse requires an account of the relationship between emotion and the agent’s capacity for choice. Does intense emotion incapacitate choosing capacity by definition? Moore realises that intense emotion cannot always be the enemy of an agent’s choosing capacity, not least because such an approach would give excusatory effect to emotions whatever their moral quality.⁹¹ He thus accepts that certain emotions are part of choice, rather than

⁸⁶ *R v Johnson (Dean)*, *ibidem*.

⁸⁷ And it should be noted that where English sentencing is concerned, a governing principle is one of proportionality in light of offence seriousness, and offence seriousness is governed in part by the wrongdoer’s mental relationship to relevant criminal harm and the significance of that relationship for his culpability: see the Criminal Justice Act 2003, s 143. It is submitted that this mental relationship can only act as a factor in offence seriousness when the accused has a capacity for normative understanding as outlined above by Duff, see text related to n 66 above.

⁸⁸ See *Placing Blame*, above n 5, Chapter 13 ‘Choice, Character and Excuse’.

⁸⁹ Moore uses the term excuse somewhat more widely than its use in this article: he means any factor that reduces blame sufficiently to be acknowledged by the criminal law. In general terms his notion of defective capacity tracks exemptions and diminished opportunity tracks excuses, though nothing really turns on this classificatory scheme.

⁹⁰ For example, Lord Lane, whilst defining the English law of duress in the leading case of *Graham* [1982] 1 All ER 801, at 806 refers to the defendant having “good cause to fear” death or serious physical injury. Also the common law definition of provocation under English law focused, in its subjective limb, on the emotions experienced by the accused: see the reference to “passion” in the definition of loss of self-control given by Devlin J in *Duffy* [1949] 1 All ER 932: see n 113 below. Admittedly the new statutory ‘loss of control’ defence created and defined by the Coroners and Justice Act 2009 (see n 52 above) makes no reference to emotion as the term “loss of self-control” used by the Act is left undefined by the Act, but amongst the statutorily defined qualifying triggers for losing self-control is “fear of serious violence” (s 55(3)): for more discussion of this defence see n 52, n 60 and n 72 above and n 112 below and accompanying text. As for the US, the Model Penal Code’s partial defence to murder, adopted in many US states, refers to the concept of “extreme...emotional disturbance” as a ground for reducing murder to manslaughter: see section 210.3 (1) (b).

⁹¹ Above n 5, p 556.

its enemy.⁹² The challenge for his theory is to articulate when emotion incapacitates choice purely in terms of the process of practical reasoning itself, in order to distinguish meritorious cases from unmeritorious. Moore hints at the possibility that emotions incapacitate choice when disconnected from judgment or when they influence action directly, but expresses reservations about these possibilities and in any event leaves them unexplored.⁹³ It is submitted that this question of who deserves an excuse in the context of emotion cannot be resolved without reference to the agent's capacity for moral evaluation and an assessment of the moral quality of his emotion and its expression in action. This insight no doubt explains those features of formulations of the provocation and duress defences that focus on the moral quality of the agent's emotions and behaviour in light of those emotions. In turn, if the agent under circumstances of provocation or duress is to be evaluated in light of the moral quality of the emotion, in part or in full, and the actions that flow from it, this presupposes an agent who has the capacity to appreciate the normative significance of his emotions and actions, and who can offer up an explanation for his actions that is derived from the moral nature of his situation. Thus, the moral excuses of duress and provocation do, and their criminal law formulations should, presuppose agents who have a much richer normative capacity than that envisioned by choice theory.⁹⁴

Perhaps aware of this difficulty for his theory, Moore seeks some refuge in the diminished opportunity conceptualisation of duress.⁹⁵ According to this conceptualisation, opportunities are unfairly diminished when the agent acts to prevent a substantial evil, though his actions fall short of being justified. However, this conceptualisation still requires more than the raw capacity to choose otherwise; opportunities for choice are unfairly diminished because the morally discerning agent only considers his opportunities diminished in certain kinds of morally difficult situations; in turn, the agent can only be blamed for failing to meet our expectations, for example assaulting another when threatened with minor harm, because he has made a meaningful normative error. Thus the various formulations of the duress defence, with their requirements for a serious threat,⁹⁶ still presuppose an agent with a much richer normative capacity than that envisioned by choice theory.

And choice theory leaves one fundamental question unanswered, as Huigens points out: "[Moore] fails to tell us what it is about practical reason... that makes its engagement a necessary condition of inculcation."⁹⁷ The answer, as argued above, is that choices can only be the subject of blame when they are normatively meaningful. When this is not the case, allocating blame is an empty gesture. It is not enough to rely on the intuition that

⁹² *Ibidem*, pp 558-559.

⁹³ *Ibidem*, p 560.

⁹⁴ And despite the plausibility of choice theory in light of the current interpretation of the M'Naghten rules, it fails to account for the defence of infancy: an 8 year old can understand that certain conduct is criminal, but his lack of normative capacity accounts for the existence of the defence in the criminal law.

⁹⁵ Above n 5, p 560.

⁹⁶ Above n 90. See also (for the US) the Model Penal Code section 2.09.: "It is an affirmative defense that the actor engaged in the conduct charged to constitute an offense because he was coerced to do so by the use of, or a threat to use, unlawful force against the person or the person of another, which a person of reasonable firmness in his situation would have been unable to resist."

⁹⁷ 'Virtue and Inculcation' above n 80, at 1442.

successful practical reasoning makes the agent morally responsible. It is when that practical reasoning is embedded in a wider capacity for moral understanding that the agent becomes morally responsible.

The above reveals that choice theory is perhaps most plausible as a theory of criminal responsibility that expressly seeks a measure of dislocation between criminal responsibility and moral culpability. It is arguable that, since moral capacity is a subtle and contestable concept, defences based upon it might introduce harmful indeterminacy in the criminal law. Perhaps confining exempting defences to, as Arenella summarises it, the very crazy and the very young enables the criminal law to achieve the societal functions of deterrence and incapacitation whilst at the same giving limited acknowledgment to the idea that the criminal law should not punish the blameless. Such a theory would require re-interpreting those normative features of the defences of duress and provocation not in terms of culpability, but in terms distilling those who are truly dangerous for society from those who are not. What is interesting about this approach is that it takes choice theory very close to Lacey's character theory. It would also require a sentencing regime that did not undermine its broad approach at conviction. But, most importantly, it would no longer be a theory that took the connection between criminal responsibility and moral culpability seriously, a connection that is the *raison d'être* of choice theory.

The Rational Capacity Theory versus the Self-Control Theory

Two capacity theories remain which I shall term the rational capacity theory and the self-control theory.⁹⁸ Both theories incorporate the requirement for moral understanding in accordance with the Duff definition quoted above, and do so for the same reasons; the disagreement between them concerns whether or not a capacity for self-control is also a necessary condition of moral responsibility. The rational capacity theory says no,⁹⁹ the self-control theory says yes. In this part of the article I will argue that the self-control theory is the superior theory of moral and criminal responsibility.

According to the rational capacity theory, if wrongdoing springs from the agent's intelligible understanding of the pertinent factual and normative features of his situation, he is morally responsible without more. Unlike character theory, which only considers the agent morally responsible for his wrongdoing when it is linked to his character, the rational capacity theory adopts a broader definition of the agent. According to the rational capacity theory, all his cognitive and emotional features, however transient or hidden from consciousness, constitute the agent.¹⁰⁰ The result is that wrongdoing that is in the

⁹⁸ The term 'rational capacity theory' is my creation and describes the ideas found in the writings of various authors. Its various components are to be found in Eugene Schlossberger *Moral Responsibility and Persons* (Philadelphia: Temple University Press, 1992); Robert M Adams 'Involuntary Sins' (1985) 94 *Philosophical Review* 3; Simester 'Can Negligence be Culpable?' in Horder (ed) *Oxford Essays in Jurisprudence* 4th Series (Oxford: Oxford University Press, 2000) p 85 and Gardner 'The Gist of Excuses' (1998) 1 *Buffalo Criminal Law Review* 575, reprinted in *Offences and Defences* above n 48, p 121.

⁹⁹ Schlossberger *ibidem*, p 41 and p 67, Adams *ibidem*, pp 12-13; Simester *ibidem*; Gardner *ibidem*, at 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. I disagree: see my 'Capacity, Moral Responsibility and the Criminal Law' [2007] *Denning Law Journal* 33.

¹⁰⁰ Schlossberger, *ibidem* pp 34 and 44 and generally Adams, *ibidem*.

required sense linked to those features is wrongdoing for which the agent is morally responsible.¹⁰¹ Thus, according to the rational capacity theory, the agent cannot identify himself exclusively with his character, and thereby deny moral responsibility for wrongdoing that is out of character.¹⁰² Furthermore, he cannot identify himself exclusively with his better judgment, and thereby deny moral responsibility for wrongdoing that is the product of weakness of will.¹⁰³ This broad definition of the agent, combined with the requirement that action be rationally intelligible, means the rational capacity theory confines exemptions to two categories: those that deny any link between the agent and his wrongdoing because the wrongdoing was involuntary (that is to say the link between the theory's broad conception of the agent and his 'action' was severed in some way) and those that deny that the agent is capable of rational action *tout court*.¹⁰⁴ The self-control theory accepts that involuntariness and a fundamental incapacity for rational action should exempt, and does so for the same reasons. However, it holds that there is more to moral responsibility than a capacity for intelligible action. According to the self-control theory, the agent must also possess a distinct and active capacity to control his conduct in light of his conception of the good. The capacity is active because it enables the agent to put his settled higher order values into practical effect whatever the desires, emotions, urges etc he is experiencing; that is to say it enables the agent to ignore desires, emotions, urges etc when they prompt conduct incompatible his conception of the good.¹⁰⁵ It is worth noting that this requirement for self-control is something the rational capacity theory explicitly rejects.¹⁰⁶ As a result, the self-control theory does not confine exemptions to involuntariness and fundamental denials of rational capacity: it includes a further category of exemption for those experiencing extreme difficulties with self-control. Some examples of such difficulties will be discussed below. The difference in approach of the two theories to moral responsibility is reflected in each theory's account of excuses. The rational capacity's exclusive focus on rational motives means it conceives all excuses as based on the moral quality of the agent's intelligible motives for his wrongdoing.¹⁰⁷ Lacey summarises this approach as follows: "The focus, crucially, is upon the quality of the attitude manifested *in the defendant's conduct, evaluated in the light of (a generous interpretation of) the context in which it occurred* [author's emphasis]".¹⁰⁸ The self-control theory accepts that certain excuses may function

¹⁰¹ Schlossberger, *ibidem* p 41, p 44 and p 67, Adams, *ibidem*, and Simester 'Can Negligence be Culpable' above n 98, p 98.

¹⁰² Schlossberger, *ibid* p 44. The rational capacity theory's conception of the agent is therefore all embracing: as such, it does not suffer from the defects that flow from treating out of character wrongdoing as a denial of moral responsibility.

¹⁰³ Schlossberger, *ibidem* p 41 and p 67.

¹⁰⁴ Gardner above n 98, at 587-590.

¹⁰⁵ I am indebted to the work of Wallace for this notion of an active capacity for self-control: see, for example, 'Addiction as Defect of the Will: Some Philosophical Reflections' (1999) 18 *Law and Philosophy* 621 and 'Three Conceptions of Rational Agency' (1999) 2 *Ethical Theory and Moral Practice* 217.

¹⁰⁶ Schlossberger above n 98, p 41 and p 67; Adams above n 98, at 12-13, Simester above n 98, at 98 and Gardner above n 98, at 581-582, 584-585 and 589.

¹⁰⁷ The rational capacity 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 98 generally.

¹⁰⁸ '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

purely by dint of the moral quality of the agent's motives; indeed, the defence of duress is arguably best accounted for by such a rationale. However, in light of its acknowledgement of the relevance of self-control, the self-control theory includes a further category of excuse where, despite an agent's possession of an intact capacity for self-control, he nevertheless fails to exercise self-control under certain extreme circumstances. We shall see immediately below that it conceives of the defence of provocation in such terms.

Why should the self-control theory be favoured over the rational capacity theory? There are two reasons: first, the rational capacity theory cannot account for the excusatory power of the defence of provocation;¹⁰⁹ secondly, the rational capacity theory gives an incomplete account of exemptions, an account completed by the self-control theory. In turn, the self-control theory, by completing a picture left incomplete by the rational capacity theory, exposes a failure by the criminal law to track moral culpability. I will deal with each of these reasons in turn.

We have seen that rational capacity theorists concentrate on the agent's capacity for rational motivation, with excuses allocated in accordance with the moral quality of that motivation. In their accounts of the rationale of the various formulations of the defence of provocation, rational capacity theorists have, as a result, located that defence's excusatory power exclusively in the agent's good reasons for being angry.¹¹⁰ The problem is that this purely evaluative approach to provocation is insufficient to explain the excusatory power of the defence in any form. Unlike situations of self-defence and duress, where the agent acts to protect autonomy in the face of some pressing need, actions taken under provocation are backward looking, motivated by the agent's sense of injury.¹¹¹ In light of the agent's failure to accord the correct moral weight to the value of the provoker's life and the importance of the lawful resolution of disputes, this backward looking quality, in and of itself, is fatal to an excuse, whatever the moral quality of the agent's anger. Something else requires articulation if we are to understand why we should excuse in such cases. The rational capacity theory, given its exclusive focus on motivation, cannot

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), on the basis that 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 rational capacity theory.

¹⁰⁹ Where English law is concerned, the common law defence of provocation has been replaced with a wholly statutory partial defence of 'loss of control', created and defined by the Coroners and Justice Act 2009, see above n 52. A necessary component of this defence is a loss of self-control but there is no definition of this term within the 2009 Act. Furthermore this loss of self-control must be due to certain qualifying triggers defined by s 55(3)-(5). The qualifying trigger defined at s 55(3) does refer to the emotion of fear, in the form of a "fear of serious violence", but it is submitted that this will be interpreted as merely the anticipation of such violence.

¹¹⁰ See Gardner 'The Gist of Excuses' above n 98. See also Dan M Kahan and Martha C Nussbaum 'Two Conceptions of Emotion in Criminal Law' (1996) 96 *Columbia Law Review* 269, analysing the American formulations of the defence, which bear a close relation to the former common law definition of the provocation defence under English law. For similar arguments to those of Kahan and Martha C Nussbaum, see Victoria Nourse 'Passion's Progress: Modern Law Reform and the Provocation Defense' (1997) 106 *Yale Law Journal* 1331.

¹¹¹ Admittedly, the new 'loss of control' defence's qualifying triggers incorporate both backward looking and forward looking elements: see above n 52, n 60 and n 109.

supply that reason. The self-control theory can: intense and justified anger triggered by the provocation makes the exercise of self-control very difficult. This difficulty, in combination with the moral quality of the agent's anger, allows us to overlook, albeit partially, the moral failures involved in the act of killing. The self-control theory therefore completes a picture left incomplete by the rational capacity theory. It holds that, despite the failure of self-control, in view of the gravity of the provocation *and* the difficulties intense anger causes for the exercise of self-control, the agent has exercised his capacity for self-control to a normatively fixed standard, and, as a result, can be partially excused.

Admittedly the new statutory defence of 'loss of control' that has replaced the common law defence of provocation makes no reference to emotional states, but to a "loss of self-control" and certain qualifying triggers.¹¹² Arguably this new defence therefore reflects the purely evaluative approach of the rational capacity theory, whereas the common law provocation defence's frequent explications in terms of emotional disturbance were more in tune with the self-control theory.¹¹³ However, it is submitted that the courts should interpret the new defence's reference to "loss of self-control" to incorporate a requirement for emotional disturbance.

The second reason the self-control theory is to be favoured over the rational capacity theory is because the rational capacity theory gives an incomplete account of exemptions. According to rational capacity theorists, once the agent is the author of his wrongdoing, exemption is only available if the agent is fundamentally incapable of rational action. This approach can be illustrated with John Gardner's account of the distinction between the common law defence of provocation and the statutory defence of diminished responsibility where 'provoked' agents are concerned, as his argument is informed by the rational capacity theory.¹¹⁴ Gardner argues that all excuses are cases of what he terms "indirect rational explanation."¹¹⁵ 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, in keeping with the rational capacity's exclusive focus on evaluation of motivation, provocation excuses when the wrongdoer can point to intelligible and good reasons for her anger. In turn, according to Gardner, the way that provocation differs from diminished 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".¹¹⁶ In other words, the agent is fundamentally incapable

¹¹² See above n 52. S 54 of the Coroners and Justice Act 2009, which defines the partial defence, is entitled 'loss of control', but the text of the other sections that flesh out the defence make frequent reference to 'loss of self-control'. The definitions of the qualifying triggers are given by s 55(3) to (5).

¹¹³ For English law see *R v Duffy* above n 90, where Lord Goddard, in the Court of Criminal Appeal, quoted with approval the following words of Devlin J at first instance: provocation requires in its subjective limb "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". *Duffy* was approved in *R v Ibrams* (1981) 74 *Criminal Appeal Reports* 154, *R v Thornton (Sara Elizabeth)* (No 1) [1992] 1 All ER 306 and *Richens* [1993] 4 All ER 877.

¹¹⁴ It should be noted that the defence of diminished responsibility as set out in s 2 of the Homicide Act 1957 has been amended by the Coroners and Justice Act 2009 s 52(1), though the changes introduced by the 2009 Act do not matter to the above discussion.

¹¹⁵ See 'The Mark of Responsibility' above n 68, at 158.

¹¹⁶ Gardner 'No Provocation Without Responsibility: A Reply to Mackay and Mitchell' (2004) *Criminal Law Review* 213, at 214-215.

of intelligible action and as a result a partial exemption, and hence partial defence, is appropriate.

The problem with Gardner's vision is that it does not accommodate certain phenomena that should exculpate but do not fall into either his conception of exemption or excuse. I will illustrate this with involuntary intoxication and the case of *Kingston*.¹¹⁷ 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.¹¹⁸ 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. According to Gardner's vision, Kingston merits neither an exemption nor an excuse.¹¹⁹ This is because, as Gardner might put it, his rationality "goes all the way down": it is perfectly intelligible, when confronted with an object of sexual attraction, to take advantage of that opportunity.¹²⁰ As such, according to Gardner, his moral responsibility is intact. Nor, given the fact that his motives were morally reprehensible, does he deserve an excuse in Gardner's eyes. Nevertheless it is submitted that such an agent does merit exculpation and only the self-control theory can explain why.¹²¹ It does not conceive of that exculpation in excusatory form; this is because, unlike provocation where an ordinary person has lost self-control in the face of grave provocation and has therefore satisfied certain normative expectations, Kingston's act of morally reprehensible abuse entirely fails to meet normative expectations. However, an exemption is appropriate in light of the fact that a core feature of Kingston's agency was severely compromised, in the form of the adverse effect on his capacity for self-control of his non-voluntary consumption of the soporific drugs. These adverse effects freed Kingston from the standards of self-control we expect from non-intoxicated agents.

The self-control theory, accepting that a capacity for self-control is a necessary condition of moral responsibility, can therefore account for the exculpatory effect of involuntary intoxication by holding that when the effects of the intoxicant are severe enough, the agent's capacity for self-control has been sufficiently compromised that he deserves an exemption.¹²² And the facts of *Kingston* do not exhaust the range of factors that can impact on self-control. These include various phenomena grouped under the heading of impulse control disorders, including intermittent explosive disorder and the various forms of addiction. The rational capacity theory, due to its insistence, once the agent is the

¹¹⁷ [1995] 2 AC 355, (1994) 99 Cr App R 286.

¹¹⁸ "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 'Making Excuses' above n 53, p 131, n 3.

¹¹⁹ Some might think that the disinhibition created by alcohol/drugs functions in the same manner as automatism, by breaking all links between the agent and the harm brought about by his bodily movements. However, the global view of the agent adopted by the rational capacity theory (see above ns 100 to 103 and related text) means the agent remains the author of his wrongdoing in such circumstances. The expression *in vino veritas* springs to mind.

¹²⁰ 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.

¹²¹ Tadros has endeavoured to use his alienation theory to explain why we might exculpate in such cases of intoxication but his explanation is in my opinion not convincing, see 'The Characters of Excuse' above n 50, pp 298-302. Limitations of space prevent me from setting out why.

¹²² This may be partial: matters depend on the intensity of the effects of the impairing factor.

author, as it sees it, of his wrongdoing, that exemptions are confined to the fundamentally irrational or non-rational, is not in position to account for the exculpatory effect of these phenomena. It therefore represents an impoverished vision of the types of incapacity that can negate moral responsibility.

The above reveals that there is a gap in the lexicon of criminal law defences. If criminal responsibility is to track moral culpability comprehensively it should be reformed to create defences for those whose capacity for self-control has been severely compromised. The exact form such a defence should take is no doubt a technical matter, involving questions such as 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.¹²³ The administration of such a defence also raises issues over the allocation of the burden of proof, and also important forensic questions. However, suffice to say for present purposes that, if the criminal law is to reflect accurately the moral culpability of agents, the self-control theory reveals a need for a defence based on radical impairment of the agent's capacity for self-control.

Character, Capacity and Negligence

Crimes of negligence are numerous and increasing in number, embracing a wide range of circumstances where criminal harms are unjustifiably risked and on certain occasions also caused. Generally speaking the genesis of criminal negligence falls into two categories: cognitive and behavioural. These categories are not mutually exclusive. With cognitive negligence, the negligence is constituted by the agent's faulty belief management prior to and/or during conduct, with the result that he goes about his conduct in ignorance of an unjustified risk; with behavioural negligence, the imperfect performance of a task creates an unjustified risk of which the agent is not aware. In both cases, there is the judgment that the agent has fallen below standards of care to be expected of the reasonable person. The legitimacy of blame in such cases is dependent on whether the agent is culpable for the failure to reach this standard.

Character theorists have sought to account for the culpability of negligence by linking the above failures to the agent's possession of a stable attitude of indifference. There is no doubt that the both categories of negligence can be the product of such an attitude: with the first category, the attitude explaining why the person formed the (incorrect/incomplete) beliefs that he did; in the second category the attitude explaining why, for example, the agent is not concentrating whilst performing a task. But there remains the fundamental problem that even if the negligence is born of such an attitude, it can still be a one-off rather than reflecting anything stable within the agent's character. Thus all the objections to character theory outlined above apply in the context of negligence.

Capacity theory locates culpability for negligence in the agent's possession of general capacities that he fails to exercise on occasions of negligence, those being in general terms his cognitive capacities for accurate belief formation and general dexterity. This is

¹²³ It is worth noting that the amendments to the diminished responsibility defence made by the Coroners and Justice Act 2009 make specific reference to the agent experiencing substantial impairment of his ability to exercise self-control as a way of grounding the defence: see s 52 (1) and (1A), amending s 2 of the Homicide Act 1957.

what Hart termed the culpability of ‘unexercised capacity’.¹²⁴ Beyond this general statement, however, it becomes difficult to locate the exact source of the agent’s culpability. If the failure to exercise existing capacities is due to a defective attitude, as outlined above, then blame can focus on the underlying defective attitude.¹²⁵ In turn, if the failure is due to the absence of a relevant cognitive or physical capacity, for example a physical disability or stupidity, then the agent should be exempt from blame because of his inability to avoid wrongdoing. The difficulties exist because there are instances of failure to reach standards that do not have their origin in permanent or transient lack of care, yet the agent’s capacities nevertheless appear intact. The moment’s inattention on the road, the random failure of memory, the misreading of a label because it has letters in common with another word... The examples are numerous. There is no failure of attitude to ground the blame in such cases. Where does culpability lie on such occasions? The self-control theory offers the following answer to this question. It locates blame in the agent’s capacity to exercise his capacities and thereby avoid the negligent wrongdoing. The danger with this approach is that it is vulnerable to the argument that, at the relevant time, the agent could not in fact avoid the wrongdoing, that is to say those who ‘do not’ in fact ‘cannot’, and hence lack the relevant capacity at the operative moment.¹²⁶ This and similar observations have led some to conclude that negligence is not culpable.¹²⁷

In fact this problem does not end with negligence, because it is ultimately the problem of determinism. This is where the character and rational capacity theories have a distinct advantage over the self-control theory. By locating culpability in a defective attitude, both theories need not care whether that attitude was something over which the defendant had control or not: the attitude can be blamed *simpliciter*. Both theories are therefore compatibilist and though this is not the place to address the determinism debate, it is necessary to acknowledge that the self-control theory, rooting moral culpability as it does in the ability to control conduct in light of a conception of the good, is vulnerable to the challenge of determinism in a way that the character and rational capacity theories are not.

CONCLUSION

¹²⁴ See *Punishment and Responsibility*, above n 6, pp 145-57.

¹²⁵ The success of such an approach perhaps depends on the agent being able to monitor and correct his own defective attitudes when they run the risk of leading to negligent behaviour.

¹²⁶ Perhaps the source of blame in such cases is based on the distinction between the ‘can-general’ and ‘can-particular’: the agent is to blame because, generally, he can get the task right: see T Honoré ‘Responsibility and Luck’ (1988) *Law Quarterly Review* 530 at 550-552.

¹²⁷ In particular Alexander in *Crime and Culpability* above n 67, Chapter III: ‘Negligence’. Alexander and his co-authors believe negligence is not culpable for three reasons: first, it does not manifest ‘insufficient concern’, the concept that grounds blame in Alexander’s eyes, and which, as he sees it, is exclusively constituted by the conscious taking of an unjustified risk with the victim’s legally protected interests; secondly, because attempts to find insufficient concern or some other morally culpable attitude in some instances of negligence fail according to Alexander or collapse negligence into recklessness (see pp 71-81); thirdly, and finally, because attempts to construct a reasonable person test as the benchmark standard against which negligence is measured are arbitrary (see pp 81-85). I do not find all of these arguments convincing but limitations of space prevent me articulating my reservations here; see also Moore, *Placing Blame*, above n 5, pp 418-419.

Despite the spectre of determinism, the self-control theory remains, in both its descriptive and evaluative power, the most insightful and comprehensive theory of criminal responsibility. Character theory, despite its initial plausibility, simply fails on closer examination, not least because its plausibility can only be maintained by emptying it of all content. The other versions of capacity theory are either incomplete, as is the case with choice theory and the rational capacity theory, or make claims that cannot ultimately be justified, as is the case with the good moral motivation theory. Thus the self-control theory, by including both the requirements of moral capacity and self-control, thereby incorporates all the elements necessary to explain and justify a theory of criminal responsibility rooted in moral culpability.