Passion's Puzzle

Stephen P. Garvey
Cornell Law School, spg3@cornell.edu

Follow this and additional works at: http://scholarship.law.cornell.edu/facpub
Part of the Criminal Law Commons

Recommended Citation
http://scholarship.law.cornell.edu/facpub/255

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Publications by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.
Passion's Puzzle

Stephen P. Garvey*

ABSTRACT: The puzzle of the provocation defense, otherwise known as the "heat of passion" defense, is to figure out how, if at all, each of the basic elements or features of the doctrine can be explained in a coherent and normatively attractive fashion. None of the prevailing theories of provocation can solve this puzzle. These theories either fail to explain one or more of the doctrine's basic elements, or else end up committing the state to a decidedly illiberal course of action: punishing citizens not only for what they do (for their actions), but for who they are (for their characters).

The theory advanced here purports to offer a solution. According to this theory, called the akrasia theory, the basic elements of the defense work in concert to achieve the normatively attractive goal of sorting actors who kill in defiance of the law (and who should therefore be convicted of murder) from those who kill in a moment of culpable ignorance of law or weakness of will (and who should therefore be convicted of the lesser crime of manslaughter). Insofar as this theory justifies and so defends the basic contours of existing provocation doctrine, it challenges those who view the doctrine, in some or all of its formulations, as a pernicious presence in the criminal law.

* Professor of Law, Cornell Law School. I thank John Blume, Joshua Dressler, Sheri Johnson, Trevor Morrison, and Steven Shiffrin for their comments and insights.
INTRODUCTION ........................................................................................................... 1679

I. THE PUZZLE'S PIECES ......................................................................................... 1686

II. FAILED SOLUTIONS .......................................................................................... 1692
    A. PARTIAL-JUSTIFICATION THEORIES ......................................................... 1693
        1. Worthy Motives ..................................................................................... 1693
        2. Disproportionate Response ................................................................. 1695
    B. PARTIAL-EXCUSE THEORIES ................................................................. 1698
        1. Incapacity ............................................................................................ 1700
            a. Total Incapacity ............................................................................. 1701
            b. Partial Incapacity ....................................................................... 1703
        2. Unfair Opportunity ............................................................................. 1706
    C. NEITHER PARTIAL JUSTIFICATION NOR PARTIAL EXCUSE ................... 1717
        1. Evaluative Emotions ......................................................................... 1718
        2. Warranted Excuse ........................................................................... 1723

III. A SOLUTION TO THE PUZZLE ........................................................................... 1726
    A. PROVOCATION AS AKRASIA ..................................................................... 1727
        1. Two Forms of Criminal Culpability .................................................. 1727
        2. Putting the Pieces Together ............................................................... 1731
            a. Reasonable Loss of Self-Control .................................................. 1731
            b. Partiality ....................................................................................... 1732
            c. Passion and Adequate Provocation ............................................ 1732
    B. DIMINISHED CAPACITY .......................................................................... 1738

CONCLUSION ............................................................................................................. 1745
INTRODUCTION

The following vignettes are based on facts drawn from reported cases:

- Claudia Brenner and Rebecca Wight were hiking along the Appalachian Trail in the mountains of south central Pennsylvania. The couple stopped for the night and made camp. They were "engag[ed] in lesbian lovemaking" when a "mountain man" named Stephen Carr happened upon them. Armed with a .22-caliber rifle, Carr shot the two women, wounding Brenner and killing Wight. Carr said he had become enraged at the sight of the "'show' put on by the women, including their nakedness, their hugging and kissing and their oral sex."  

- Alma was the sixteen-year-old daughter of Sealous Grugin. Another of Grugin's daughters, Louella, was married to Jeff Hadley, who had "carried [her] away." Grugin disliked Hadley's "bad habits," and "bad blood" soon formed between the two men. About a year later, Hadley raped Alma, who reluctantly disclosed the assault to her father the following month. Armed with a shotgun, Grugin went to Hadley's home. He intended to "take charge of [Hadley] until [he] could get an officer." Approaching Hadley, Grugin asked, "Jeff, whatever possessed you to rape Alma, my daughter?" Hadley replied, "I will do as I damn please about it," whereupon Grugin shot and killed him in "hot blood."  

- Morgan Smith was a carpenter. One evening he "received a visit from his old friend James McCullagh." Smith and McCullagh were "both alcoholics and spent the evening in drinking and recrimination." Smith's grievances against McCullagh "went back many years." Most recently, Smith believed McCullagh...
had stolen some of his carpentry tools, sold them, and used the money to "buy drink." McCullagh denied the charge, and as he repeated the denial, Smith became furious. With the "row in full swing," Smith grabbed a knife and stabbed McCullagh several times, killing him.

In each of these cases the defendant claimed to have killed in the "heat of passion," or what is more commonly today called provocation. The doctrine of provocation has been part of the law for a very long time. According to one statement of its common-law formulation, an intentional homicide, which would otherwise constitute the crime of murder, is reduced to voluntary manslaughter if the defendant killed upon adequate provocation in the sudden heat of passion. But while the doctrine of provocation is well-established, the theory of provocation is not. Provocation is a partial defense to murder, but why?

Traditional theories portray provocation as a species of justification or excuse. According to these theories, provocation reduces murder to

14. Id.
15. Id.
17. See JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW § 31.07, at 527 (3d ed. 2001). Prior to the enactment of section 3 of the 1957 Homicide Act, the “classic direction” given to English juries in cases alleging provocation was set forth in Regina v. Duffy, [1949] 1 All E.R. 932, 933 (Crim. App.). Current English law governing provocation is set forth in section 3 of the 1957 Homicide Act. See Homicide Act, 1957, 5 & 6 Eliz. 2, c. 11, § 3 (Eng.) (“Where on a charge of murder there is evidence on which the jury can find that the person charged was provoked . . . to lose his self-control, the question whether the provocation was enough to make a reasonable man do as he did shall be left to be determined by the jury.”); see also ANDREW ASHWORTH, PRINCIPLES OF CRIMINAL LAW § 7.4(b), at 268-80 (4th ed. 2003) (describing contemporary developments in English law); A.P. SIMESTER & G.R. SULLIVAN, CRIMINAL LAW: THEORY AND DOCTRINE § 10.5, at 342-60 (2d ed. 2003) (same). For general descriptions of the defense in American law, see, for example, DRESSLER, supra, § 31.07, at 527-33; WAYNE R. LAFAVE, CRIMINAL LAW § 7.10, at 703-17 (3d ed. 2000); 1 PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 102, at 479-91 (1984). The Model Penal Code (MPC) mitigates murder to manslaughter if the killing was "committed under the influence of extreme mental or emotional disturbance for which there is reasonable explanation or excuse." MODEL PENAL CODE § 210.3(1)(b) (1985). A comparative study of Indian, English and Australian law on provocation can be found in STANLEY YEO, UNRESTRAINED KILLINGS AND THE LAW: PROVOCATION AND EXCESSIVE SELF-DEFENCE IN INDIA, ENGLAND AND AUSTRALIA (1998).
18. An actor who satisfies the conditions of a full justification is conventionally portrayed as one who has done nothing the criminal law considers wrongful. See, e.g., GEORGE FLETCHER, RETHINKING CRIMINAL LAW § 10.1, at 759 (1978) (“A justification speaks to the rightness of the act . . . .”); B. Sharon Byrd, Wrongdoing and Attribution: Implications Beyond the Justification-Excuse Distinction, 33 WAYNE L. REV. 1289, 1290 (1987) (“A justification focuses on the wrongfulness of an actor’s conduct. It presupposes the violation of a legal or moral norm, but provides that this
manslaughter because the killing was either justified or excused, though only partially.\textsuperscript{19} Partial-justification theories say that an adequately provoked actor’s killing is wrong, but not as wrong as it would have been if the actor had been unprovoked or inadequately provoked.\textsuperscript{20} Partial-excuse theories say that an adequately provoked actor’s killing is just as wrong as an unprovoked or inadequately provoked actor’s killing, but the adequately provoked actor is somehow less responsible or less blameworthy than his unprovoked or inadequately provoked counterpart.\textsuperscript{21} Some theories try to

violation was the right thing to do under the circumstances."); Joshua Dressler, Justifications and Excuses: A Brief Review of the Concepts and the Literature, 33 WAYNE L. REV. 1155, 1161 (1987) (“In its simplest form . . . justified conduct is conduct that is a good thing, or the right or sensible thing, or a permissible thing to do.”) (internal quotations and citation omitted); Heidi M. Hurd, Justification and Excuse, Wrongdoing and Culpability, 74 NOTRE DAME L. REV. 1351, 1558 (1999) (“Justified actions should be conceived of as right actions . . . .”); cf. Mitchell N. Berman, Justification and Excuse, Law and Morality, 53 DUKE L.J. 1, 4 (2003) (“A justified action is not criminal.” (emphasis omitted)).

An actor who satisfies the conditions of a full excuse is conventionally portrayed as one who has done something wrongful, but he has done so in or under circumstances rendering him blameless. See, e.g., Fletcher, supra, at 759 (“[A]n excuse [speaks] to whether the actor is accountable for a concededly wrongful act.”); Byrd, supra, at 1290 (“An excuse . . . focuses on the blame we may fairly attribute to the actor.”); Dressler, supra, at 1162–63 (“An excuse is in the nature of a claim that although the actor has harmed society, she should not be blamed or punished for causing that harm.”); Hurd, supra, at 1558 (“[E]xcused actions should be conceived of as wrong actions done nonculpably.”); cf. Berman, supra, at 4 (“[A]n excused defendant has committed a crime but is not punishable.” (emphasis omitted)).

19. Some jurisdictions recognize partial defenses to murder other than provocation. Most notable among these defenses are diminished capacity and excessive or imperfect self-defense. See generally PARTIAL EXCUSES TO MURDER pts. I & III (Stanley Meng Heong Yeo ed., 1990). For commentary on the nature of partial excuses or defenses, see, for example, Douglas N. Husak, Partial Defenses, 11 CAN. J.L. & JURISPRUDENCE 167, 177 (1998), proposing a “unifying hypothesis” according to which a “given circumstance qualifies as a partial defense if and only if it bears a given relation to a complete justification or excuse”; Martin Wasik, Partial Excuses in the Criminal Law, 45 MOD. L. REV. 516, 533 (1982), concluding among other things that the “traditional arguments against partial excuses must surely be reconsidered.” See also Suzanne Uniacke, What Are Partial Excuses to Murder?, in PARTIAL EXCUSES TO MURDER, supra note 19, at 1, 14–15 (“A partial legal excuse . . . recognizes that . . . the accused’s degree of culpability is sufficiently reduced for it to be appropriate to convict him or her of a lesser offense.”).

20. Finbarr McAuley is the scholar whose name is probably most commonly associated with the partial-justification theory. See Finbarr McAuley, Anticipating the Past: The Defence of Provocation in Irish Law, 50 MOD. L. REV. 133, 156 (1987) [hereinafter McAuley, Anticipating] (“[I]t is submitted that a defendant who can show that he killed in the face of substantial provocation should, on this ground alone, be entitled to the defence.” (emphasis added)); accord Finbarr McAuley, Provocation: Partial Justification, Not Partial Excuse, in PARTIAL EXCUSES TO MURDER, supra note 19, at 19, 29. Alan Norrie has also recently argued that modern English provocation law, though purporting to be a partial excuse, functions in reality as a “surreptitious[,]” partial justification. Alan Norrie, The Structure of Provocation, 54 CURRENT LEGAL PROBS. 307, 338 (2001).

21. Joshua Dressler is the most prominent defender of a partial-excuse theory. Although Dressler has modified his theory over the years, the basic idea has remained more or less the same. A provoked actor is partially excused, according to Dressler, if and because he “barely could” control his desire to kill in the face of the provocation. See Joshua Dressler, Provocation:
combine elements of justification and excuse, usually saying that the adequately provoked actor’s anger may have been justified, but that the killing itself is merely excused, and then only partially.\textsuperscript{22}


\textsuperscript{22} For various arguments to this effect, see ASHWORTH, supra note 17, § 6.5(b), at 233 ("The element of wrongdoing by the victim (or another) might therefore be combined with the excusable element of loss of self-control to provide grounds for a (partial) defence."); HORDER, supra note 16, at 156 ("[T]he defence of provocation is an excuse, although it involves significant elements of moral justification . . . "); SAMUEL H. PILSBURY, JUDGING EVIL: \textit{RETHINKING THE LAW OF MURDER AND MANSLAUGHTER} 141 (1993) ("We grant mitigation when a person has been provoked because . . . the person had good reason for strong anger, and . . . in experiencing such anger, the person’s moral restraints were temporarily strained to the breaking point."); A.J. Ashworth, \textit{The Doctrine of Provocation}, 35 CAMBRIDGE L.J. 292, 307 (1976) ("[T]he doctrine of provocation as a qualified defence rests just as much on notions of justification as upon the excusing element of loss of self-control."); Marcia Baron, \textit{Killing in the Heat of Passion, in SETTING THE MORAL COMPASS: ESSAYS BY WOMEN PHILOSOPHERS} 355, 366-69 (Cheshire Calhoun ed., 2004) (suggesting how an understanding of "provocation as a hybrid of justification and excuse" might work); Timothy Macklem & John Gardner, \textit{Provocation and Pluralism}, 64 MOD. L. REV. 815, 819 (2001) (arguing that provocation is an excuse but "in making an excuse one relies on the fact that one’s unjustified action was taken on the strength of a justified belief or attitude or emotion"); Uma Narayan & Andrew von Hirsch, \textit{Three Conceptions of Provocation}, CRIM. JUST. ETHICS, Winter/Spring 1996, at 15, 18-19 (proposing a
The theory advanced here portrays provocation as neither justification nor excuse, partial or otherwise. Rather than being squeezed into either category, provocation is better understood as a doctrine through which the law tries to distinguish killings committed in defiance of the law (murder) from those committed in a moment of culpable ignorance of law or weakness of will (manslaughter). Provocation therefore distinguishes between two different forms or modes of criminal culpability. Consequently, portraying provocation as an imperfect justification or excuse is a misportrayal.

An actor who knows that the law forbids killing, but nonetheless wholeheartedly chooses to kill, defies the law. In contrast, an actor who kills in the heat of passion upon adequate provocation violates the law, but he does not defy it, either because he honestly and momentarily (though unreasonably) believes the law allows him to kill (a case of culpable ignorance of law), or because, though he realizes the law permits no such thing, and though believing it best to comply and wanting to comply, he fails

"moral conflict" model of provocation in which the "provoked actor's diminished culpability [is explained] in terms of the conflict in moral sentiments evoked by the victim's wrongful conduct"); Samuel H. Pillsbury, Emotional Justice: Moralizing the Passions of Criminal Punishment, 74 CORNELL L. REV. 655, 678 (1989) ("The distinction [between a valid case of provocation and an invalid one] lies not in differences in the feeling of sudden anger but in the reason for such anger."); Andrew von Hirsch and Nils Jareborg have "propose[d] that the common law requirements [of provocation] be replaced by two separate tests, reflecting the two conceptions of impaired volition and resentment, respectively." Andrew von Hirsch & Nils Jareborg, Provocation and Culpability, in RESPONSIBILITY, CHARACTER, AND THE EMOTIONS 241, 253 (Ferdinand Schoeman ed., 1987). In other words, provocation should be split into a partial-excite defense and a partial-justification defense, with the "doctrinal particulars" adjusted accordingly. Id. at 255.

23. Others have also claimed provocation is neither an excuse nor a justification as those terms are commonly understood. See, e.g., Dan Kahan & Martha C. Nussbaum, Two Conceptions of Emotion in Criminal Law, 96 COLUM. L. REV. 269, 320 (1996) (stating that "justification" and 'excuse' fail to explain the [provocation] doctrine...[as those] concepts [are] currently understood"); Victoria Nourse, Passion's Progress: Modern Law Reform and the Provocation Defense, 106 YALE L.J. 1331, 1397-99 (1997) (stating that neither partial-justification theories nor partial-excite theories can explain all the doctrinal requirements of provocation but ultimately characterizing her proposed theory as an excuse-based theory). I suggest that the Kahan-Nussbaum and Nourse theories of provocation are best understood as partial-justification theories, see infra pp. 1717-22 (discussing Kahan-Nussbaum); infra pp. 1723-26 (discussing Nourse), although they can also be interpreted as excuse-based theories in which the adequate-provocation requirement functions as a forfeiture rule. See infra pp. 1709-17 (explaining how the adequate-provocation requirement can function as a forfeiture rule).

24. For more on these two forms of culpability, see Stephen P. Garvey, Two Kinds of Criminal Wrongs, 5 PUNISHMENT & SOCY 279, 280-90 (2003). See also RONALD D. MILO, IMMORALITY 234 (1984) (distinguishing "six main types of immoral behavior"). Milo's focus is on forms of immoral behavior, while mine is on forms of illegal behavior. The form of immorality Milo calls "preferential wickedness" corresponds mutatis mutandis to what I call "defiance," while that which I call "akratic" encompasses and corresponds mutatis mutandis to what Milo calls "moral negligence" and "moral weakness." Id.
to do so (a case of culpable weakness of will). Either way, the provoked killer’s culpability consists in something other than willful defiance of the law.

Instead of defiance, the provoked killer’s culpability consists in his failure to control the desire to kill, a desire resulting from provocation and the heat of passion, such that at the moment he acts, his desire to kill leads him to believe (erroneously and unreasonably) that the law allows him to kill; or, knowing the law does not allow him to kill and believing he should conform to its demand not to kill, he nonetheless succumbs to his lethal desire. The Greek word for such failures is *akrasia*. On this view, provocation is neither an excuse nor a justification, at least as those terms are commonly understood in criminal-law scholarship. Rather, provocation is a way of distinguishing those whose action manifests one form of culpability (defiance) from those whose action manifests another (akrasia). Provocation sorts culpable killers into the appropriate category.

25. For present purposes no distinction is drawn between “weakness of will” and “akrasia,” which are both understood to involve an act freely and intentionally done against the actor’s better judgment. See, e.g., ALFRED R. MELE, IRRATIONALITY: AN ESSAY ON AKRASIA, SELF-DECEPTION, AND SELF-CONTROL 7 (1987). According to Mele:

An action $A$ is a strict incontinent action if and only if it is performed intentionally and freely and, at the time at which it is performed, its agent consciously holds a judgment to the effect that there is a good and sufficient reason for his not performing an $A$ at that time.

Id. For an alternative analysis according to which weakness of the will consists in an actor’s failure to act on an intention formed consistent with his better judgment, see Richard Holton, *Intention and Weakness of Will*, 96 J. PHILO. 241, 241 (1999).


27. Anthony Duff has recently written:

[I]t cannot plausibly be argued that one whose commission of a crime exhibited weakness of will rather than true Aristotelian vice should therefore be acquitted, or should be convicted only of a lesser offense. Whether the offense is serious . . . or relatively minor . . ., we can imagine it being committed either through weakness of will by one who gives in to a temptation that he knows, in some sense, he should resist, or through vice by one who is wholeheartedly involved in the wrong he does: but both commit . . . the same offense.

R.A. Duff, *Virtue, Vice, and Criminal Liability: Do We Want an Aristotelian Criminal Law?*, 6 BUFF. CRIM. L. REV. 147, 168–69 (2002); see also Stephen J. Morse, *Reason, Results, and Criminal Responsibility*, 2004 U. ILL. L. REV. 363, 388 ("As a positive matter . . . current criminal law grades neither crimes nor punishments according to the agent’s wholeheartedness."). Setting aside possible questions about the relationship between “true Aristotelian vice” and my conception of defiance, the analysis advanced here maintains that in at least one pocket of the law—provocation—the law does convict of a lesser offense those whose crime exhibited weakness of will. Indeed, Duff goes onto say, with hesitation but seemingly with approval, that “it might be argued that the distinction [between vice and weakness of will] does become relevant at the stage of sentencing.” Duff, *supra*, at 169; see also Morse, *supra*, at 388 (suggesting that one might think “as a normative matter that [the criminal law] should” grade crimes and punishments “according to [an] agent’s wholeheartedness”). The basis for Duff’s recognition of the
The argument unfolds in three parts. Part I describes the essential elements or features of provocation. These elements form the pieces, so to speak, of provocation's puzzle. A successful theory of provocation must be able to assemble these pieces in a coherent and normatively attractive manner. Part II argues that none of the prevailing theories of provocation passes this test. Some fail to explain all of the doctrine's elements. Others succeed in doing so only at the cost of punishing some provoked actors, not for what they have done, but for who they are. In other words, these latter theories punish some actors, not for the character of their actions, but for the content of their characters. Such an illiberal use of state power renders these theories decidedly unappealing. Part III sets forth the theory of provocation as akrasia and argues that, unlike its competitors, it manages to pull all the pieces of provocation's puzzle together into a normatively attractive, or at least acceptable, whole. Part III also discusses the relationship between provocation so conceived and the kindred doctrine of diminished capacity.

My purpose here is to try to construct a theory to explain and justify, and so defend, what I take to be provocation's essential elements. Accordingly, I avoid taking sides in ongoing debates over whether, for instance, the doctrinal specifics of the defense should be formulated more along the lines of the Model Penal Code (MPC) or more along the lines of the common law.
provocation, reduced to its essentials, should be understood or portrayed, and whether the law should recognize it as a defense at all. Moreover, a defense of provocation’s essentials is needed now more than ever, for the number of scholars misportraying provocation, as well as the number calling for its outright abolition, appears on the rise. My hope is to stem the tide.

I. THE PUZZLE’S PIECES

An actor who intentionally kills another, and who, in the absence of a valid excuse or justification, would therefore be guilty of murder, will instead applicable to all crimes. See infra note 187 (discussing proposals for such a generic partial excuse).

31. The misportrayals I have in mind, both of which are discussed in some detail below, are twofold: first, efforts to portray provocation as a partial justification, see infra text accompanying notes 56-68; and second, efforts to portray it as a partial excuse in which the adequate-provocation requirement functions as a forfeiture rule, see infra text accompanying notes 92-99. The former portrayal, whatever its normative merits or demerits, cannot explain provocation doctrine as we know it, while the latter is normatively unattractive insofar as it ends up punishing inadequately provoked actors not for what they have done (for their acts), but for who they are (for their characters).

32. Such calls can be found in HORDER, supra note 16, at 194-95 (arguing that provocation should be abolished because, among other things, the desire for retaliatory suffering on which it is based is “ethically ‘defeasible’”); Matthew Goode, The Abolition of Provocation, in PARTIAL EXCUSES TO MURDER, supra note 19, at 55 (concluding that the “arguments for abolishing the defence are, taken overall, much stronger than the arguments for retaining it, unless one is in a jurisdiction in which murder penalties remain inflexible”); Jeremy Horder, Autonomy, Provocation and Duress, 1992 CRIM. L. REV. 706, 712 (urging elimination of provocation as a “confession and avoidance” defense (like duress) and emphasizing “the very real potential of provocation to be pleaded, just as intoxication is pleaded, as a straightforward denial of specific intent, in all crimes (not just murder) where such an intent must be proven”); Kyron Huigens, Homicide in Aretaic Terms, 6 BUFF. CRIM. L. REV. 97, 134 (2002) (stating that provocation “ought to be abolished as pernicious and unnecessary”); Celia Wells, Provocation: The Case for Abolition, in RETHINKING ENGLISH HOMICIDE LAW 85, 85 (Andrew Ashworth & Barry Mitchell eds., 2000) (concluding that provocation should be abolished as a defense because it “is bound to encourage and exaggerate a view of human behaviour which is sexist, homophobic, and racist”); cf. Stephen J. Morse, Diminished Rationality, Diminished Responsibility, 1 OHIO ST. J. CRIM. L. 289, 290 (2003) (“I... believe, buttressed by more recent feminist analysis, that traditional provocation/passion doctrine is unwise.”). The Law Reform Commission of the Australian state of Victoria has also recently recommended the abolition of provocation. See VICTORIA LAW REFORM COMM’N, DEFENCES TO HOMICIDE: FINAL REPORT xlv (2004) (“The partial defence of provocation should be abolished. Relevant circumstances of the offence, including provocation, should be taken into account at sentencing as they currently are for other offences.”), available at http://www.lawreform.vic.gov.au/CA256902000FE154/Lookup/Homicide_Final_Report/$file/FinalReport.pdf.

33. See, e.g., DRESSLER, supra note 17, § 31.03[A], at 506. Provocation is usually raised against a charge of intent-to-kill murder, though it can be and has been recognized in many (but apparently not all) jurisdictions as a partial defense to other forms of murder, such as intent to inflict serious bodily harm and depraved-heart murder. See, e.g., id., § 31.07[A], at 527. LAFAVE, supra note 17, § 7.10, at 709-04. Jeremy Horder has suggested that “[i]n so far as a particular (kind of) provoked retaliation is a spontaneous response to sudden feeling, like bowing one’s head in shame or putting one’s hands to one’s head in despair, it may in some
be found guilty of manslaughter if he loses self-control in the face of adequate provocation and kills while in the heat of passion, provided his loss of self-control was reasonable. So understood, the provocation doctrine has four basic elements or features:

1. The provocation to which the actor responds must have been adequate;

2. The killing must have occurred while the actor was in the "heat of passion";

3. The actor must have lost self-control, and his loss of self-control must have been reasonable, such that a reasonable person in the actor's situation would likewise have lost control and killed; and

4. Such cases be regarded as intentional but not as specifically intended." Horder, supra note 32, at 714. Such a claim presupposes that an action can coherently be described as intentional but not intended. Assuming such a description is indeed coherent, cases of provocation meeting it would be cases in which the defendant should be able to assert a failure-of-proof defense based on lack of mens rea.

34. The passion or emotion (I treat the terms equivalently) implicated in cases of provocation is usually thought to be anger or resentment, inasmuch as anger or resentment is the emotion produced or experienced in response to provocation. Those who favor excuse-based theories of provocation are apt to be more ecumenical with respect to the range of passions on which an actor can rely in support of a plea of provocation. See, e.g., Dressler, Sexual Advances, supra note 21, at 746 n.108 ("Although courts usually talk about passion as if anger were the only emotion that the law recognizes, any intense emotion will suffice." (citation omitted)). Nonetheless, provocation might fairly be limited to cases involving anger insofar as the desire attending anger, unlike the desire attending other emotions, is the desire to retaliate against or punish a person in response to his culpable wrongdoing. See, e.g., AARON BEN-ZE'EV, THE SUBLTETY OF EMOTIONS 381-82 (2000); ROBERT C. ROBERTS, EMOTIONS: AN ESSAY IN AID OF MORAL PSYCHOLOGY 204 (2003); ROBERT C. SOLOMON, THE PASSIONS 288-89 (1976); James R. Averill, Anger, in NEBRASKA SYMPOSIUM ON MOTIVATION 1978 71-73 (Richard A. Dienstbier ed., 1979). In this regard, however, one must keep in mind that one emotion can lead to another, as when shame leads to anger, or when envy leads to shame, which leads finally to anger. See, e.g., JON ELSTER, ALCHEMIES OF THE MIND: RATIONALITY AND EMOTIONS 328-31 (1999) (noting the envy-shame-anger dynamic in the context of explaining "that the place of the emotions in the explanation of behavior is very complex").

35. The "reasonable loss of self-control" requirement—if taken at face value—turns out to be an insuperable obstacle to any coherent theory of provocation. The problem can be simply stated: a reasonable person never loses his self-control to the point of using lethal violence, no matter what the provocation. See, e.g., FLETCHER, supra note 18, § 4.2.1, at 247 ("[I]n the context of provocation, the reasonable person is hardly at home .... [A]s everyone is prepared to admit, the reasonable person does not kill at all, even under provocation."); Glanville Williams, Provocation and the Reasonable Man, 1954 CRIM. L. REV. 740, 742 ("[H]ow can it be admitted that that paragon of virtue, the reasonable man, gives way to provocation?"). If so, then provocation should never be a defense to murder. See, e.g., SIMESTER & SULLIVAN, supra note 17, § 10.5(b)(iii), at 348 ("[Provocation] should rarely succeed: for how often, surely, would the reasonable person intentionally kill, in peacetime, except in self-defence or to protect the lives of others?"). Alternatively, if an actor's loss of self-control was reasonable, then provocation should be a full defense, and not simply a partial one. See, e.g., id. ("[I]f it really were true that a
(4) The defense afforded to an actor who satisfies the conditions identified in (1) through (3) is a partial defense. It mitigates murder to manslaughter, but does not provide a full or complete defense.85

Different jurisdictions spell out the first and third of these requirements in different ways. At common law, for example, an alleged provocation was

reasonable person in D’s situation would also have killed V, the appropriate verdict should be an acquittal.”); Uniacke, supra note 19, at 13. According to Uniacke:

[P]rovocation is inadequately defined as a partial excuse if its conditions require that a reasonable person would have lost self control in the circumstances. (If this strong condition is met, provocation should be a complete excuse because the accused cannot reasonably be expected not to have lost self-control. For this reason, the condition increasingly emphasized by the courts—that an ordinary person could have lost self-control in the circumstances—more accurately captures the notion of a wrongful act which is blameworthy, but sufficiently humanly understandable in the circumstances to be partially excused.)

Id.

Accordingly, the reasonable loss of self-control requirement cannot be taken at face value. The literature suggests four strategies for dealing with this problem. First, it can be dealt with through elimination: the reasonable loss of self-control requirement can simply be eliminated and treated as no part of the doctrine. Second, it can be dealt with through subjectification: the reasonableness of the actor’s loss of self-control can be evaluated in comparison to a “subjectified” reasonable person standard, which taken to its logical conclusion amounts to a roundabout way of eliminating the requirement, since the issue is no longer whether the actor’s loss of self-control was reasonable, but rather (and simply) whether the actor did in fact lose self-control. Third, it can be dealt with through consolidation: the reasonableness of an actor’s loss of self-control can be defined exclusively in terms of the adequacy of the provocation, such that the “reasonableness” element of the reasonable loss of self-control requirement is again effectively eliminated, since the requirement would no longer have any meaning or function independent of the adequate-provocation requirement. Fourth, and perhaps most commonly, it can be dealt with through substitution: the “ordinary” person can be substituted for the “reasonable” person; or what “a” reasonable person “might” have done for what “the” reasonable person “would” have done; or what it would not have been “unreasonable” for a reasonable person to have done for what it would have been “reasonable” for him to have done. I ultimately endorse a version of the substitution strategy. See infra pp. 1726-45.

36. Provocation is partial not only because its effect is limited to mitigating murder to manslaughter. It is also partial because its scope is limited: it applies to no crime other than murder. This limitation is fairly subject to criticism. See, e.g., Ashworth, supra note 17, § 6.5(e), at 234 (“[If there is a case for ... a broadly drawn defence based on emotional or mental disturbance [as in the MPC’s manslaughter provision], is it right that it should be available only in murder cases[?]”); Stephen J. Morse, Excusing and the New Excuse Defenses: A Legal and Conceptual Review, in 23 Crime and Justice: A Review of Research 529, 598 (Michael Tonry ed., 1998) ("Why should partial excuse mitigation not be available to an arsonist acting in the heat of passion . . . , in extreme mental or emotional disturbance, or with substantially impaired mental responsibility?"); see also B. Sharon Byrd, On Getting the Reasonable Person Out of the Courtroom, 2 Ohio St. J. Crim. L. 571, 577 (2005) (developing an approach to provocation that "generalizes the provocation defense so it can apply to all criminal offenses"); Wasik, supra note 19, at 523 (observing that "[i]n the case of provocation . . . there is a great deal of evidence in [some] jurisdictions of the partial excuse operating to affect liability . . . beyond the scope of homicide").
inadequate as a matter of law if it fell outside certain narrow categories of adequacy, and was adequate as a matter of law if it fell within them. Moreover, assuming loss of self-control played any role at common law, an actor who killed in the face of adequate provocation was presumed to have done so because he lost control. Any such loss of control was also presumed to have been reasonable, provided the lapse of time between the provocation and the killing was insufficient to have permitted a reasonable person to cool off and regain self-control.

In contrast, the MPC mitigates murder to manslaughter if the killing took place "under the influence of extreme mental or emotional disturbance for which there is reasonable explanation or excuse." The

37. The classic statement of the four common-law categories of adequate provocation is contained in Lord Holt's opinion in Regina v. Mawngride, 84 E.R. 1107, 1114-15 (Q.B. 1707). See also MODEL PENAL CODE § 210.3 cmt. 5(a), at 57-58 (1980) (discussing traditional categories of adequate provocation); HORDER, supra note 16, at 30-39 (discussing the historical development of the traditional categories in English law).

38. An uncontroversial statement of the common law rule is impossible. Cf. LAW COMM'N, supra note 16, § 3.43, at 38 ("Examination of the development of the defence [in English common law] reveals that before 1957 it had no consistent moral basis."). With that reality in mind, Jeremy Horder maintains that the early English common-law cases and commentaries (up until the mid-eighteenth century or so) made "no mention whatsoever ... of the modern notion of 'loss of self-control.'" HORDER, supra note 16, at 42. Instead, "judges and commentators used only such terms as the 'heating' or the 'stirring' of the blood to express the subjective condition" of the defense. See id. at 42 n.115. The element of loss of self-control entered English common law sometime thereafter, such that the common-law categories were no longer understood as circumstances under which the provoked actor was partially justified in killing, but rather as evidence supporting the actor's claimed loss of self-control. See id. at 87-89; see also Ashworth, supra note 22, at 255. According to Ashworth:

"[T]he [common law] categories of provocation were seen as indicating the relative gravity of the provocation and the consequent probability that it was an understandable weakness of self-control rather than wickedness of mind which caused the offence.... [But i]t was implicit in the categories ... that proof of sudden loss of self-control alone was insufficient to establish the defence of provocation.

Id.

Loss of self-control is an explicit element of section 3 of the 1957 Homicide Act. See Homicide Act, 1957, 5 & 6 Eliz. 2, c. 11, § 3 (Eng.) ("Where on a charge of murder there is evidence on which the jury can find that the person charged was provoked ... to lose his self-control, the question whether the provocation was enough to make a reasonable man do as he did shall be left to be determined by the jury ....").

39. MODEL PENAL CODE § 210.3(1)(b) (1985). The MPC provides that the "reasonableness of such explanation or excuse shall be determined from the viewpoint of a person in the actor's situation under the circumstances as he believes them to be," id., thus inviting some unspecified degree of subjectification or individualization of the reasonable person standard. See MODEL PENAL CODE § 210.3 cmt. 5(a), at 62-63 (1980). Moreover, because it covers extreme mental disturbance, as well as extreme emotional disturbance, and because it requires the "reasonableness" of the actor's explanation or excuse to be assessed from the viewpoint of a person in the actor's "situation," the MPC formulation is intentionally designed to encompass cases sounding not only in provocation but also in diminished capacity. See id. § 210.3 cmt. 5(b), at 72. I discuss the relationship between provocation and diminished capacity infra pp. 1738-44.
MPC formulation eliminates the common law's categories of adequate provocation, requiring instead a "reasonable explanation or excuse" for the disturbance under whose influence the actor killed. Like the common law, the MPC's manslaughter provision makes no explicit reference to loss of self-control, but the associated Commentary nonetheless puts loss of self-control front and center. "In the end," according to the Commentary, "the question is whether the actor's loss of self-control can be understood in terms that arouse sympathy in the ordinary citizen."\(^{41}\)

The common law's "heat of passion" defense and the MPC's "extreme mental or emotional disturbance" defense are the usual reference points for academic commentary on provocation.\(^{42}\) The common-law formulation tends to be more "objective," while the MPC's tends to be more "subjective."\(^{43}\) The common-law formulation prefers rules (the categories of adequate provocation), while the MPC's prefers standards (reasonableness).\(^{44}\) The common-law formulation, so it might be claimed, is best understood as a partial justification, while the MPC's is best understood as a partial excuse.\(^{45}\) Some commentators express sympathy for the common-law formulation (or something akin to it),\(^{46}\) while others express sympathy for the MPC's (or something akin to it).\(^{47}\)

Without elaborating on these differences, and without meaning to minimize their importance, the focus here will be on what I take to be the

\(^{40}\) Indeed, although provocation can constitute a "reasonable excuse or explanation," the text of the MPC eliminates any explicit reference to provocation. See MODEL PENAL CODE § 210.3 cmt. 5(a), at 60–61 (1980) ("[T]he Code does not require that the actor's emotional distress arise from some... provocative act perpetrated upon him by the deceased."). I thank Joshua Dressler for emphasizing this point.

\(^{41}\) Id. at 63.


\(^{43}\) See, e.g., DRESSLER, supra note 17, § 31.07[B][2][b][ii], at 531 (stating that the MPC formulation is more subjective insofar as it requires the reasonableness of the explanation or excuse for the actor's disturbance to be assessed from the viewpoint of a person "in the actor's situation").

\(^{44}\) See, e.g., id. § 31.07[B][2][b][i], at 529 ("The rigid common law categories of 'adequate provocation' have given way in many states [under the influence of the MPC] to the view that the issue of what constitutes adequate provocation should be left to the jury to decide" based on general notions of reasonableness.).

\(^{45}\) See, e.g., id. § 31.07[C][2][a]–[b], at 535–36 (discussing the fit between various features of the doctrine and partial-justification and partial-excuse theories).

\(^{46}\) See, e.g., Kahan & Nussbaum, supra note 23, at 323 ("If the theme of the common law manslaughter cases is 'virtuous rage,' the theme of the Model Penal Code is 'pathology.'").

\(^{47}\) See, e.g., Dressler, Difficult Subject, supra note 21, at 997 ("I have noted certain flaws in the Model Penal Code manslaughter provision[, but d]espite my criticisms, the Code's overall effort strikes me as positive.").
doctrine's basic or essential elements. For convenience's sake, these will be referred to as: (1) the adequate-provocation requirement; (2) the passion requirement; (3) the reasonable loss of self-control requirement; and (4) the partiality requirement. The first three requirements constitute the core elements of the defense itself, while the fourth describes its legal effect, which is to mitigate, not to exonerate.48

This statement of the provocation doctrine's basic elements is bound to elicit controversy. Provocation enjoys no canonical definition.49 Those who see provocation as a partial justification might, for example, say that loss of self-control, reasonable or otherwise, is not really part of the doctrine, or if it is, it should not be, since the adequacy of the provocation to which the actor responds is all that should matter.50 In contrast, those who see provocation as a partial excuse might say that adequate provocation doesn't really

48. This statement of the doctrine makes no explicit reference to any "suddenness" requirement, according to which the killing must occur in the sudden heat of passion. Nor does it make explicit reference to any "cooling off" requirement, according to which an actor is denied the defense if a reasonable person would have cooled off and regained his self-control in the lapse of time between the provocation and the killing. Nor, finally, does it make explicit reference to any "proportionality" or "reasonable relationship" requirement, according to which an actor would be denied the defense if the manner in which he killed was out of proportion to the provocation to which he was responding. For example, an actor would be denied the defense if the manner of the killing was especially brutal in relation to an otherwise adequate provocation.

In my view, each of these requirements can plausibly be understood to provide evidence supporting the actor's claim to have killed only because he lost self-control. So, for example, if a killing is committed suddenly (i.e., immediately upon witnessing the provocation), the more likely the actor killed because, as a matter of fact, he lost self-control. See, e.g., SIMESTER & SULLIVAN, supra note 17, § 10.5(ii)(b), at 346. ("The need for a "sudden and temporary loss of self-control" indicates that there must be something in the nature of a spontaneous response to the provocation if it is to be claimed that D lost self-control.".) Likewise, the less time that elapses between the provocation and the killing (i.e., the less time to cool off), the more likely the actor killed because, as a matter of fact, he lost self-control. See, e.g., id. ("[A]ny lapse of time between the last provocative incident and D's reaction will undermine D's claim to have lost control."). Similarly, if the killing is done in a manner proportionate to the provocation, the more likely the actor killed only because, as a matter of fact, he lost self-control, whereas a killing done in a manner disproportionate to the provocation might suggest that the actor did not even try to exercise self-control. The proportionality requirement, which traces its origins in English law to Mancini v. Dir. of Pub. Prosecutions [1942] A.C. 1, 9 (1941), "has not explicitly [been] implanted in American law, Dressler, Heat of Passion, supra note 21, at 450, and moreover, is "no longer a rule of law" in England, although "traces of it remain." ASHWORTH, supra note 17, § 7.4(b)(iii), at 278.

49. See, e.g., FLETCHER, supra note 18, § 4.2, at 242 ("[P]rovocation does not enjoy a standard definition."). Dressler helpfully recounts the various "imprecise description[s] of the elements of the defense" given by the common-law courts in Dressler, Heat of Passion, supra note 21, at 430-44.

50. See, e.g., McAuley, Anticipating, supra note 20, at 136 ("[T]he defendant who can show that he killed in the face of substantial provocation should, on this ground alone, be entitled to the defence." (emphasis added)).
belong, since loss of self-control is all that counts.\textsuperscript{51} Or, one might say that adequate provocation and reasonable loss of self-control are not really independent elements, since an alleged provocation is adequate if and only if it would cause a reasonable person to lose his self-control, and conversely, a reasonable person would only lose control in the face of adequate provocation.\textsuperscript{52} Still, insofar as none of the elements can be reduced to the others, and insofar as each has some plausible claim for inclusion as an element of the defense, a theory capable of accounting for all of them is, \textit{ceteris paribus}, better than one incapable of doing so.

Passion's puzzle is to figure out how, if at all, each of the elements or features of the doctrine of provocation can be brought together into a coherent and normatively attractive whole, and a theory of provocation can be seen as an effort to solve that puzzle. Many such theories have been offered, but none has managed to find a satisfying solution.

II. FAILED SOLUTIONS

Contemporary academic debate over the theory of provocation is usually framed as a debate between those who argue provocation is (or should be understood as) a partial justification,\textsuperscript{53} and those who argue it is (or should be understood as) a partial excuse,\textsuperscript{54} with most contemporary commentators treating it in one way or another as a partial excuse.\textsuperscript{55} Two partial-justification theories, and four partial-excuse theories can be found in, or extracted from, the literature.\textsuperscript{56} In the end, none provides the solution sought.

\textsuperscript{51} See, e.g., Dressier, \textit{Provocation}, supra note 21, at 475 n.46 ("I would . . . not require evidence of wrongful conduct as a prerequisite to finding manslaughter.").

\textsuperscript{52} See, e.g., \textit{ASHWORTH}, \textit{supra} note 17, \textsection 7.4(b)(ii), at 271 ("[N]ot every act of provocation . . . should be allowed as the basis of this qualified defence, but only those serious enough to unbalance the behaviour of a person with reasonable self-control."); \textit{LAFAVE}, \textit{supra} note 17, \textsection 7.10, at 705 ("What is really meant by 'reasonable provocation' is provocation which causes a reasonable man to lose his normal self-control . . . ."); \textit{Ashworth}, \textit{supra} note 22, at 298 ("[T]he reasonable man . . . functions [in English law] as a means of assessing the gravity [or adequacy] of the provocation.").

\textsuperscript{53} See sources cited supra note 20.

\textsuperscript{54} See sources cited supra note 21.

\textsuperscript{55} See, e.g., \textit{Berman}, \textit{supra} note 18, at 73 ("According to the majority view, provocation is a partial excuse . . . .").

\textsuperscript{56} The discussion of partial-excuse theories presented below presupposes that the criminal law is an institutional mechanism for the expression of blame, with excuses serving to deflect such blame under appropriate circumstances. See, e.g., Sanford H. Kadish, \textit{Excusing Crime}, 75 Cal. L. Rev. 257, 264 (1987) ("To blame a person is to express a moral criticism, and if the person's action does not deserve criticism, blaming him is a kind of falsehood . . . . It is this feature of our everyday moral practices that lies behind the law's excuses."). For reasons others have ably set forth, I make no effort to explain provocation in utilitarian terms. See \textit{Dressler, Difficult Subject}, \textit{supra} note 21, at 963-66; \textit{Dressler, Heat of Passion}, \textit{supra} note 21, at 434 n.136.
A. PARTIAL-JUSTIFICATION THEORIES

According to partial-justification theories, the provocation doctrine mitigates murder to manslaughter because an actor who intentionally kills in the face of adequate provocation is, though guilty of a crime, guilty of a lesser crime than an actor who intentionally kills in the face of no provocation or in the face of inadequate provocation. Such theories try to explain why an adequately provoked actor's crime is a lesser crime. Two such theories can be identified. The first focuses on the nature of the adequately provoked actor's motives (the worthy-motive theory); the second focuses on the nature of his action in response to the provocation (the disproportionate-response theory).

1. Worthy Motives

The law generally permits a private actor to kill only in self-defense or in defense of others. It does not permit an actor to kill in response to

57. The concept of a "partial justification" is puzzling. A justified action is usually understood to mean an action one is (at least) permitted to do, and any particular action can be described as either permissible or impermissible. But it makes no sense to say a particular action is "partially permitted." The logic of permission is all-or-nothing. See Uniacke, supra note 19, at 16 n.8 (arguing that the "concept of a partial justification for a particular act or offence [does not] make[] sense" (emphasis omitted)). So-called partial-justification theories of provocation should therefore be described as lesser wrong theories. The adequately provoked actor's killing is less wrongful than, but in the eyes of the law just as impermissible as, any other unjustified killing.

58. See, e.g., MODEL PENAL CODE §§ 3.04(1), 3.05(1) (1985). The MPC also permits an actor to kill when doing so is the lesser of two evils, id. § 3.02(1), and under limited circumstances when necessary for the protection of property. Id. § 3.06(3)(d). In addition, the MPC allows an actor to claim duress as a defense to murder. Id. § 2.09(1). Yet insofar as the MPC treats duress as an excuse, and not a justification, an actor who kills under duress should not be understood as having the law's permission to kill. Instead, he should be understood as having the permission of some other system of norms to which the law concedes or defers. See infra note 90 (discussing duress). Neither necessity nor duress were recognized as defenses to murder at common law. See DRESSLER, supra note 17, § 22.02, at 287-89 (describing the "parameters of the common law defense" and stating that "at least arguably the necessity defense does not apply in homicide cases"); id. § 23.04, at 905 (stating that the "common law rule . . . is that duress is not a defense to an intentional killing").
provocation, adequate or not. But an actor who kills upon adequate provocation is one who has been seriously wronged in one way or another according to some set of norms. The adequately provoked actor's anger is therefore justified or warranted. According to the worthy-motive theory, the law nonetheless continues to insist that he not express his anger with lethal violence. However, if he does kill, he does so on the basis of a motive more worthy, all else being equal, than the motive of one who kills in the absence of adequate provocation. The law therefore considers the adequately provoked killer less culpable than his unprovoked or inadequately provoked counterpart. Accordingly, he deserves to be convicted of the lesser crime of manslaughter.\footnote{Although the basis for his claim that provocation is a partial justification is not entirely clear, McAuley appears to embrace a worthy-motive theory. See McAuley, *Anticipating*, supra note 20, at 21 (noting that "the defence entails a denial that the defendant’s actions were entirely wrongful in the first place," in the sense that it "implies that the defendant was partially justified in reacting as he or she did because of the untoward conduct of his or her victim"). Signs of the worthy-motive theory can also be detected in Kahan & Nussbaum, *supra* note 23, at 313 ("[T]he law... condemn[s] acts that reflect at least some appropriate emotional motivations less severely than acts that reflect only inappropriate emotional motivations.").}

The worthy-motive theory can readily explain the adequate-provocation and partiality requirements. An adequately provoked actor is one who kills in retaliation for a serious wrong.\footnote{The adequacy of an alleged provocation can be determined according to different systems of norms. See, e.g., Fletcher, *supra* note 18, § 4.2, at 243 (social norms); Horder, *supra* note 16, at 158 (moral norms derived from Aristotelian moral theory); Kahan & Nussbaum, *supra* note 23, at 310 (social norms); Macklem & Gardner, *supra* note 22, at 827-28 (social norms related to social roles); McAuley, *Anticipating*, supra note 20, at 138 (social norms); Nourse, *supra* note 23, at 1392 (legal norms). Whatever system of norms is used, any determination as to the adequacy of an alleged provocation will be uncertain and controversial insofar as the content of the relevant norms is itself uncertain and controversial. Uncertainty and controversy will also arise when the provocation facing an actor is adequate according to one system of norms but inadequate according to another.} The adequate-provocation requirement certifies the seriousness of the wrong and thus the worthiness of the actor's retaliatory motive. The theory can also explain the partiality requirement. The adequately provoked actor's motive for killing is more worthy in comparison to the motive of an unprovoked or inadequately provoked actor, but the adequately provoked actor has nonetheless done something the law prohibits, despite his relatively more worthy motive. As such, he should still be punished, albeit less so than if he had acted on a less worthy motive.

Unfortunately, the worthy-motive theory fails to explain the passion and reasonable loss of self-control requirements. Insofar as the theory focuses on the worthiness of the provoked actor's motive for killing, it makes no difference whether the actor killed in hot blood or cold blood, nor whether at the time of the killing the actor tried to control his desire to kill or not.\footnote{See, e.g., Dressler, *Heat of Passion*, supra note 21, at 458 ("[I]f heat of passion is a partial justification, the injustice [of the decedent’s provocation], not the passion [of the actor], is the primary focus."); id. ("To be consistent, passion should not be required [on a partial-}
So long as the defendant killed for the right reasons or motives (i.e., in response to adequate provocation), neither the heat of the passion (if any) within which he acts nor his failure to exercise control over his desire to kill has any bearing on his claim to the defense. An actor who acts wholeheartedly in executing a worthy retaliatory desire to kill deserves the defense just as much as one who kills only because he fails to exercise self-control and thus succumbs to his desire. The worthy-motive theory therefore draws no distinction between the righteous vigilante and a similarly-situated actor who kills only because he "loses it."

2. Disproportionate Response

Another partial-justification theory focuses not so much on the provoked actor's motive for killing as on the nature of his response to the provocation. This theory recognizes, as does the worthy-motive theory, that the law refuses to permit an actor to resort to any violence, lethal or non-lethal, in the face of provocation, adequate or otherwise. At the same time, it also recognizes that the law itself sometimes responds to wrongdoing with violence (imprisonment), and in fact, sometimes with lethal violence...
(capital punishment). Thus, if the wrongdoing committed by a provocateur constitutes a crime, the law itself would (upon trial and conviction) punish him. It would not, of course, punish him with death, since only murder can be so punished, and the provocateur does not kill the person he provokes. He gets killed.

According to the disproportionate-response theory, an alleged provocation is adequate if and only if the provocateur’s wrong is one that the criminal law itself would punish. Consequently, an actor who kills in the face of such provocation responds with violence when the law itself would respond with violence. The adequately provoked actor’s use of violence can thus be said to be justified, but only so far as the violence he uses matches that which the law would use. The adequately provoked actor gets into trouble when and because he goes beyond the violence to which the law itself would make recourse. He kills when the law would at most imprison.

On top of that, because the state claims a monopoly on the legitimate use of violence, the adequately provoked actor who kills in response to the provocateur’s wrongdoing necessarily poaches on that authority. The provocateur may have deserved punishment, but the actor who punished him lacked the authority to do so. The state claims exclusive jurisdiction to deliver retributive punishment, and it declines to delegate that jurisdiction to private actors.

The adequately provoked actor is therefore guilty of a wrong, but only because the punishment he inflicted usurped state authority and was disproportionate to the provocateur’s crime. Thus, although he should be punished for his transgression, he should be punished less than he otherwise would have been punished had the provocateur done nothing meriting a punitive response.

The disproportionate-response theory, like the worthy-motive theory, can explain the adequate-provocation and partiality requirements. The provocation facing an actor must be adequate inasmuch as the law’s own norms, or the norms of some other system incorporated into the law or to


65. One can imagine cases in which the provocateur has killed someone other than the person he provokes. Take, for example, a case in which the defendant comes upon a provocateur who has just killed the defendant’s child. If the defendant then kills the provocateur, the defendant cannot avail himself of a defense-of-others justification because the child, as the defendant can plainly see, is already dead when the defendant arrives on the scene. However, if the killing of the child constitutes adequate provocation, the defendant may be able to plead provocation. Moreover, if the jurisdiction in question authorizes the death penalty for the provocateur’s crime, then the defendant’s resort to lethal violence in retaliation would arguably constitute a proportionate response in comparison to the law’s own response. Consequently, the disproportionate-response theory would seem committed to affording a full defense to the defendant in such a case. The only basis for denying him a full defense would be the actor’s usurpation of the state’s claimed monopoly on the legitimate use of violence.
which the law defers, would permit the use of non-lethal violence in response. If the relevant norms permitted no violence in response, then the actor's resort to lethal violence would arguably be so disproportionate as to provide no meaningful basis for mitigation. As for the partiality requirement, the law is understood to permit the use of non-lethal violence upon adequate provocation, either to avoid hypocrisy (in reference to its own use of violence) or in deference to some other system of norms, but an adequately provoked actor goes beyond the scope of this permission. He resorts to lethal violence, which is a disproportionate response. He is therefore only entitled to partial mitigation, not full exoneration.

Yet the disproportionate-response theory, again like the worthy-motive theory, cannot explain the passion or reasonable loss of self-control requirements. As the disproportionate-response theory sees it, the problem with the adequately provoked actor rests entirely on the nature of his response to the provocation. He goes too far. He might go too far because he loses control and lets passion get the better of him, but the theory requires neither passion nor loss of self-control. An actor who calmly decides

66. The disproportionate-response theory can be framed with reference to some system of norms other than those of the criminal law, such as social norms. On this view, the applicable norms, like the law's own norms, are understood to prohibit an actor from using lethal violence in response to any form of wrongdoing (except in cases of self-defense or defense-of-others). At the same time, those norms are assumed to permit the use of non-lethal violence in response to some forms of wrongdoing. Adequate provocation is therefore any wrongdoing with respect to which the applicable norms permit the use of non-lethal violence in response. The adequately provoked actor's conduct can thus be said to be justified, but only so far as he uses non-lethal violence. The problem, once again, is that the adequately provoked actor's response goes too far. He kills when the applicable norms would permit only a punch. He is therefore guilty of a wrong, but one less serious than if the provocateur had done nothing meriting a violent response.

67. Insofar as the disproportionate-response theory assumes the relevant norms would permit the use of non-lethal violence in response to adequate provocation, an actor who so responds should be entitled to a full defense, since such an actor has responded proportionately. Cf. HORDER, supra note 16, at 135. As Horder explains:

Provocation could become a complete defence... to (for example) assault, battery, or wounding where the provocation preceding it was so very grave that the offence was completely, and not merely partially, morally justified. The defence would thus operate in such cases in much the same way that duress is, in certain circumstances, a complete defence to crimes of violence other than murder.

Id. Although no jurisdiction in the United States appears to recognize such a defense, the law of some Australian states apparently does. See, e.g., ASHWORTH, supra note 17, § 6.5, at 231 n.162.

68. Cf. HORDER, supra note 16, at 135 ("Reduction of an offence to a lesser offence... is only appropriate where the defendant did exceed the mean... , but did not exceed it too greatly...."). If the applicable norms permitted the use of lethal violence, then provocation would function much like duress and should be a full defense. Provocation would function like duress insofar as duress fully excuses an actor whose violation of the criminal law, while unjustified in the eyes of the criminal law itself, is nonetheless permissible according to some other system of norms to which the law defers. See infra note 90 (explaining duress as a form of justification).
to kill a provocateur is no less entitled to a manslaughter conviction than is his impassioned counterpart who kills only because he loses control. An actor's permission to use non-lethal violence does not evaporate simply because he freely chooses to step beyond its scope and employ lethal violence.

When all is said and done, partial-justification theories cannot solve provocation's puzzle. They fail to explain the heat of passion and reasonable loss of self-control requirements. The puzzle remains unsolved.

Of course, partial-justification theories can be used to argue for changes in the existing structure of homicide law. One might, for example, reasonably believe the law should make a killing committed for a relatively worthy motive a different and lesser crime than one committed for a relatively unworthy motive, whether or not the killer acted in the heat of passion or lost self-control. An actor who intentionally kills in order to relieve the suffering of a terminally-ill friend who has pleaded for release should not, one might reasonably believe, be put into the same category as one who intentionally kills in order to humiliate or subordinate his victim. The law, one might argue, should acknowledge their different motives, convicting the former of a lesser offense and the latter of a greater one. But again, my goal here is to offer a normatively defensible account of the law as it is, not to offer advice on what it should be.

B. PARTIAL-EXCUSE THEORIES

According to the leading theory of excuse in criminal law, an actor who would otherwise be criminally liable should be excused if, through no fault of his own, he lacked either the capacity or a fair opportunity to conform his conduct to the requirements of law.\(^69\)

---

\(^69\). This theory of excuse is generally known as the "choice" theory because in simplest terms it says that an actor should be excused if he did not freely choose to break the law. See HART, supra note 28, at 152 ("What is crucial is that those whom we punish should have had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise these capacities."); see also Michael Moore, Choice, Character, and Excuse, 7 SOC. PHIL. & POLY 29, 32-40 (1990) (elaborating on Hart's general statement). The choice theory is generally contrasted with the "character" theory. See, e.g., R.A. Duff, Choice, Character, and Criminal Liability, 12 LAW & PHIL. 345, 345 (1993) (stating the usual contrast). But cf. id. at 380 (concluding that each theory "expresses, but also distorts, a significant truth about what we can properly convict and punish people for"). In fact, three very different theories of excuse—easily confused with one another—can travel under the general heading of the character theory.

According to the first version, an actor should be excused if and because his criminal action was "out of character" for him, presumably in the straightforward sense that it was unusual or atypical for him. See, e.g., FLETCHER, supra note 18, § 10.3, at 800; Michael D. Bayles, Character, Purpose, and Criminal Responsibility, 1 LAW & PHIL. 5, 7 (1982); Richard B. Brandt, A Motivational Theory of Excuses, in CRIMINAL JUSTICE: NOMOS XXVII 165, 165 (J. Ronald Pennock & John W. Chapman eds., 1985). This version of the character theory is descriptively and normatively suspect. Descriptively, the criminal law has no compunction against convicting and punishing an actor whose crime was "out of character" for him, provided he freely chose to
commit the crime. See, e.g., Moore, supra, at 51. Normatively, insofar as this version of the character theory excuses an actor if and when he acts out of character, it seems committed to saying he should be punished only when he acts in character, which seems to imply he is ultimately punished for his character, and not for his action. But a liberal state only punishes its citizens for what they do, not for who they are. As Michael Moore has put it, "no one deserves to be punished for being a poor specimen of humanity." See id. at 55. Thus, while acting out of character in the sense specified may very well ground a grant of mercy, see, for example, John Gardner, The Gist of Excuse, 1 BUFF. CRIM. L. REV. 575, 578 n.7 (1998), it should not ground an excuse.

According to the second version of the character theory, an actor should be excused if and because his actions were "out of character" inasmuch as the beliefs and desires constituting his character and on the basis of which he freely chooses to act were not authentically his beliefs and desires. The "authenticity" of an actor's beliefs and desires is widely thought to be a requirement of responsibility for actions taken on the basis of those beliefs and desires. See, e.g., JOHN MARTIN FISCHER & MARK RAVIZZA, S.J., RESPONSIBILITY AND CONTROL: A THEORY OF MORAL RESPONSIBILITY 237-38 (1998); ISHTYAQUE HAJI, MORAL APPRAISABILITY: PUZZLES, PROPOSALS, AND PERPLEXITIES 125 (1998). Although an actor can usually be presumed responsible for the beliefs and desires on the basis of which he freely chooses to act, we may in some cases be prepared to suspend this presumption. Beliefs and desires implanted through coercive indoctrination or brainwashing are the least controversial examples. A brainwashed actor acts "out of character" insofar as the beliefs and desires on which he otherwise freely acts are not really constituent elements of his character. They are more like foreign invaders. Whether this form of excuse should be recognized in the criminal law, and if so, how far it should extend has been subject to considerable debate. Does it, for example, include not only having been brainwashed, but also having come from a "rotten social background"? Compare Richard Delgado, Ascription of Criminal States of Mind: Toward a Defense Theory for the Coercively Persuaded ("Brainwashed") Defendant, 63 MINN. L. REV. 1, 10 (1978) ("In the case of the coercively persuaded defendant, it is appropriate to ask . . . whether the intent the actor possessed can properly be said to be his own."). with Joshua Dressler, Professor Delgado's "Brainwashing" Defense: Courting a Determinist Legal System, 63 MINN. L. REV. 335, 337 (1979) ("[I]t is logically impossible to frame a coercive persuasion defense that is both consistent with present criminal law . . . and . . . morally acceptable.").

At least three commentators of whom I am aware appear to suggest that provocation can be explained along the lines of this second version of the character theory. See LAWRIE REZNEK, EVIL OR ILL?: JUSTIFYING THE INSANITY DEFENSE 250 (1997) ("Provocation may excuse . . . because it induces a temporary change in moral character."); Victor Tadros, The Characters of Excuse, 21 OXFORD J. LEGAL STUD. 495, 508 (2001) ("[O]nce one is extremely angry, one's actions no longer reflect one's settled character quite as closely as they do when one is calm."); William Wilson, The Filtering Role of Crisis in the Constitution of Criminal Excuses, 17 CAN. J.L. & JURISPRUDENCE 387, 411 (2004) (submitting more or less in agreement with Tadros that passion excuses "because [it] may take us outside our constituted character and in so doing subvert our normal responsiveness to reason"). On this view, adequate provocation produces in the actor a desire to kill in retaliation. The actor then freely chooses to act on that desire. Nonetheless, the actor should be (partially) excused for acting on that desire insofar as that desire is not really a part of his character. The provocateur's actions have momentarily implanted an alien desire to kill into the actor's character, thereby "destabilizing" it. Cf. Tadros, supra, at 503. When the adequately provoked actor freely acts on the basis of this alien desire, his action is therefore temporarily "out of character."

Whatever its merits may otherwise be, this version of the character theory seems unable to explain the provocation doctrine's reasonable loss of self-control and partiality requirements. If the adequately provoked actor's desire to kill truly is an alien desire for which the actor bears no responsibility, then it would seem to make no difference whether the actor tried but failed to control the desire or not. The desire is not his in either case. But cf. id. at 507
In other words, a person should be excused if, through no fault of his own, he could not have conformed his conduct to the requirements of law (lack of capacity), or if he could, he nonetheless should not have been expected to conform, because such conformity would have imposed on him an unfair burden (lack of fair opportunity).

Either of these grounds for excuse—lack of capacity or lack of fair opportunity—can be used to develop an excuse-based theory of provocation. The idea behind any such theory is that an actor who intentionally kills in the face of adequate provocation in the heat of passion should be partially excused if and because he either could not control his desire to kill, or because demanding such control would be an unfair demand under the circumstances. Four excuse-based theories are described below, two based on lack of capacity to conform, and two based on lack of fair opportunity to conform.

1. Incapacity

Incapacity theories of provocation come in two forms. According to the total-incapacity theory, provocation excuses if and because the actor, as a result of provocation, suffers a total, though temporary, loss of capacity to control his conduct and conform to the law's demands. According to the partial-incapacity theory, the provocation confronting the actor does not cause a total loss of capacity for self-control, but it does cause a partial loss, which suffices to mitigate murder to manslaughter.

(suggesting that "loss of self-control" may be a "clumsy way of reflecting that the defendant was no longer responsive to the reasons that she was ordinarily motivated by"). Likewise, if the desire to kill truly is an alien desire, then a full excuse would seem to be in order, much like a full excuse would seem to be in order in cases involving coercive indoctrination or brainwashing, assuming such cases warrant any excuse. But cf. id. at 507 n.34 (purporting to offer an explanation for the partiality requirement).

According to the third version of the character theory, an actor is excused, not because he acts in any sense "out of character," but because, on the contrary, he lives up to some applicable norm of character. See Gardner, supra, at 578. In my view, this form of character theory is one way of fleshing out the unfair-opportunity branch of the orthodox choice theory. See Garvey, supra note 24, at 286-87. If an actor who breaks the law nonetheless acts as he is permitted to act pursuant to some other norm to which the law defers (including what might be described as norms of character), then the actor's opportunity to comply with the law was "unfair," and the law grants him a full excuse. If so, then the unfair-opportunity branch of the choice theory of excuse is really a form of justification. The actor's conduct is impermissible according to the law's own norms; consequently, the actor is not entitled to a "standard" justification, such as self-defense. Nonetheless, the actor's conduct was permissible according to some other system of norms to which the law concedes or defers. This condescension or deference takes the form of an "excuse." For more on this theory and why it cannot solve passion's puzzle, see infra pp. 1717-26.

70. Because the excuse-based theories examined below focus on an actor's failure to conform to the law, they all involve some problem with the actor's volition (his capacity and fair opportunity to conform his conduct to the law), not some problem with his cognition (his capacity and fair opportunity to know what the law requires).
a. Total Incapacity

According to the total-incapacity theory, provocation excuses an actor if and because, as a result of provocation, he experiences a desire to kill his provocateur; moreover, the actor's capacity to control this desire is at the same time rendered completely but temporarily disabled. The total-incapacity theory can readily explain the passion requirement. The heat of passion (anger) arising from the provocation constitutes, generates, or intensifies a desire to kill strong or intense enough to be completely but temporarily beyond the actor's ability to control it. Perhaps the actor cannot control the desire to kill because the desire bypasses his will altogether. The desire acts on its own, so to speak, without any intervening act of will, causing the fatal movement of the actor's body. In other words, one might say, the actor's body moved, but he did not move it. The total-incapacity theory therefore analogizes the provoked actor's bodily movement to that of a reflex, or the provoked actor himself to an automaton, or perhaps to one who is insane, acting under the influence of an "irresistible impulse" attributable to mental disease or defect.

The theory can likewise explain the reasonable loss of self-control requirement. If a desire is strong or intense enough to be beyond an actor's capacity to control it, it would a fortiori be unreasonable to expect such control. Ought implies can, and insofar as an actor is caught in the grip of a desire he truly cannot control no matter how hard he tries or were to try, such control cannot reasonably be expected of him. The theory does less well when it comes to explaining the adequate-provocation and partiality requirements. Because it focuses on the strength

71. See, e.g., Dressler, Heat of Passion, supra note 21, at 465-66 (discussing this theory).
72. The ontology of the emotions and their relationship to desire is an exceptionally complicated and controversial philosophical question. See sources cited infra notes 130-42. I will for present purposes simply assume a particular emotion can be analyzed into an associated constellation of beliefs (the emotion's cognitive element), desires or passions (its conative element), and feeling or "heat" (its affective element).
73. One might fairly claim that no desire is literally irresistible. Desires not resisted exist, but irresistible desires do not. See, e.g., Morse, supra note 36, at 357 (arguing that one can "plausib[ly]" characterize some desires as irresistible but believing it "more accurate to say" that such putative desires excuse on the basis of "some theory of psychological compulsion" or "rationality defect"). If irresistible desires do indeed turn out not to exist, so much the worse for the total-incapacity theory. My aim here, however, is to explain the total-incapacity theory, not to defend it. For an effort to explain what it means to say an impulse or desire is irresistible without making any claims as to the existence of any such impulses or desires, see Michael Smith, Irresistible Impulse, in INTENTION IN LAW AND PHILOSOPHY 37 (Ngaire Naffine et al. eds., 2001). See also ALFRED R. MELE, SPRINGS OF ACTION 97-99 (1992).
74. It would probably be more precise to say that this explanation of the reasonable loss of self-control requirement explains why it would be unreasonable to expect self-control, not why the actor's loss of self-control is reasonable. If so, then the total-incapacity theory adopts a substitution strategy for dealing with the reasonable loss of self-control requirement. See supra note 35.
or intensity of the actor's desire to kill, and not on the reason why the actor finds himself gripped with that desire in the first place, the theory cannot explain why the doctrine insists on adequate provocation. If what really matters is the existence of a desire strong enough to totally disable the actor's capacity for self-control, then any provocation generating such a desire should suffice, adequate or inadequate. Indeed, if all that matters is the desire's intensity, then the genesis of the desire is irrelevant, provided the actor is not responsible or to blame for its genesis. If so, then the total-incapacity theory not only fails to explain why the provocation facing an actor must be adequate, it also fails to explain why any provocation is required.75

Likewise, the total-incapacity theory cannot explain the partiality requirement. Insofar as the theory sees the provoked actor in the same light as one whose body moves on reflex, who is an automaton, or who is temporarily insane as a result of an irresistible impulse, it should provide him with a complete defense. An actor who causes harm on reflex or functions as an automaton is, and should be, free from any consequent criminal liability. He does not act in the sense required for criminal liability.76 Similarly, an actor who commits a crime as a result of an irresistible impulse due to mental disease or defect should be free from criminal liability, since he is (on one definition) insane.77 If the provoked actor belongs to the same family as these actors, as the total-incapacity theory implies, then the law should afford him a full defense, just as it does the rest of his family.78

75. The total-incapacity theory could explain the adequate-provocation requirement if it were to say an actor is responsible for his irresistible desire to kill when the desire arises from inadequate provocation, but not when it arises from adequate provocation. But explaining the adequate-provocation requirement in this way would commit the theory to imposing punishment on the inadequately provoked actor not for what he has done, but for who he is (for his character). See infra pp. 1710–26.

76. See DRESSLER, supra note 17, § 9.02[G][2], at 85–86 (explaining the voluntary act requirement).

77. Many jurisdictions today do not recognize "irresistible impulses" due to mental disease or defect as a basis for a plea of insanity. See 2 PAUL H. ROBINSON, CRIMINAL LAW DEFENSES § 173(e)(1), at 68 (Supp. 2004-05). The basis for rejecting such a defense appears primarily to be skepticism as to the existence of such impulses or desires. If such impulses or desire do exist, then a jurisdiction's refusal to recognize an "irresistible impulse" test of insanity must be defended on grounds of optimal error allocation. For example, insofar as it is very difficult to tell if a particular actor did indeed act on the basis of an irresistible impulse, a system without an irresistible-impulse defense will, on the one hand, never risk erroneously acquitting an actor who claimed to have suffered from such an impulse, but who in fact did not. On the other hand, such a system will risk convicting actors who did suffer from such an impulse and who should therefore be acquitted. The question then becomes one of balancing the respective costs and benefits of a system without the defense compared to one with it. See United States v. Lyons, 739 F.2d 994, 999 (5th Cir. 1984) (Rubin, J., dissenting) (arguing that the balance should be struck in favor of continued retention of a volition-based test for insanity).

78. See, e.g., ASHWORTH, supra note 17, § 6.5(a), at 231–32. As Ashworth notes:
b. Partial Incapacity

The partial-incapacity theory can be seen as an effort to redress the failure of the total-incapacity theory to explain the partiality requirement. The partial-incapacity theory focuses on the strength or intensity of the provoked actor's desire to kill arising from the provocation. But whereas the total-incapacity theory says that the provoked actor's loss of capacity to control the desire is and must be complete, the partial-incapacity theory (true to its name) says that the loss of capacity is and need only be partial. And because the provoked actor's capacity for self-control is only partially impaired, the defense to which he is entitled is only partial. The partial-incapacity theory therefore provides a facially plausible explanation for the partiality requirement.

It is not difficult to conceive of conditions of extreme rage in which a person would find it virtually impossible to control his actions. Probably there is enough in the argument that loss of self-control may occasionally negative mens rea or voluntariness to suggest that its claim to be treated as a complete defence should be considered seriously.

Id.; see also HORDER, supra note 16, at 119 ("[I]t is not clear why, if anger does indeed take the form of irresistible impulse, a complete acquittal is not the result that should of necessity follow."); Dressler, Difficult Subject, supra note 21, at 974 ("If [the provoked actor] were totally incapable, a full excuse would be defensible."); Dressler, Heat of Passion, supra note 21, at 465-66 ("If the provocation was so great that it would probably cause the ordinary, law-abiding person to wholly lose his ability to control himself, then ... the provocation should wholly, not just partially, excuse the actor."). Some cases ostensibly falling under the heading of duress might also "operate rather like insanity as an excuse." R.A. Duff, Rule-Violations and Wrongdoings, in CRIMINAL LAW THEORY: DOCTRINES OF THE GENERAL PART 47, 64 (Stephen Shute & A.P. Simester eds., 2002).

97. Joshua Dressler is the foremost proponent of the partial-incapacity theory. See sources cited supra note 21. According to Dressler, provocation is based on the incapacity branch of the choice theory of excuse, in contrast to duress, which is based on the unfair opportunity branch. See Dressler, Difficult Subject, supra note 21, at 974 n.68. Although Dressler recognizes that "[t]here are similarities between provocation and duress," id. at 975 n.68, he nonetheless ultimately rejects the analogy, preferring instead the analogy to insanity. See id. ("In duress cases, we sense that the coerced party chooses to accede to the coercer's threat; it is in a real sense a rational, albeit perhaps socially unjustifiable, choice. In contrast, with provocation, the killing is the result of an emotional explosion ..."). See id. ("[T]he provocation plea represents ... a 'barely could' [exercise self-control] claim by a criminal defendant.").

98. See Dressler, Difficult Subject, supra note 21, at 974 ("[T]he [provoked actor's] loss of self-control is not totally excusable, because the law's assumption is that the provoked party was not wholly incapable of controlling or channeling his anger."); Dressler, Heat of Passion, supra note 21, at 466 ("Heat of passion only makes sense as a partial excuse when the jury can say that the ordinarily law-abiding person would have become so sufficiently angry that he would have been unable to fully control his anger."); Dressler, Provocation, supra note 21, at 472 ("[T]he provocation plea represents ... a 'barely could' [exercise self-control] claim by a criminal defendant.").

99. But see infra pp. 1704-05 (questioning the coherence of the idea of a partial incapacity to conform one's conduct to the requirements of law).
The partial-incapacity theory can also explain the reasonable loss of self-control and passion requirements, offering much the same explanations as does the total-incapacity theory. The heat of passion is needed because it constitutes, generates, or intensifies a desire to kill in response to the provocation strong or intense enough to render the actor partially incapable of controlling it. Likewise, expecting or demanding an actor to control such a desire would be unreasonable.82 The provoked actor's failure to exercise self-control is, if not reasonable, at least not unreasonable insofar as it results from a partial disabling of his capacity for such control, which is itself the result of an intense provocation-induced desire.

In the end, the only real difference between the total- and partial-incapacity theories is the extent to which the provoked actor's capacity for self-control is said to be impaired. The total-incapacity theory, insisting as it does on total impairment, portrays the provoked killer as having acted on reflex, or as an automaton, or as being temporarily insane, killing as a result of an irresistible impulse. The partial-incapacity theory also portrays the provoked killer as temporarily insane,83 not because he kills out of an irresistible impulse, but rather because he momentarily lacks, as the MPC's definition of insanity would describe it, the "substantial capacity... to conform his conduct to the requirements of the law."84

The partial-incapacity theory is an improvement over the total-incapacity theory. Nonetheless, like the total-incapacity theory, the partial-incapacity theory cannot explain the adequate-provocation requirement. Again, the strength or intensity of the actor's desire is the linchpin of the theory. As such, it doesn't matter why the actor finds himself so gripped, provided of course that the actor himself is not at fault or to blame for being

82. As with the total-incapacity theory, the partial-incapacity theory's explanation for the reasonable loss of self-control requirement is really an explanation for why such a loss of self-control is not unreasonable, as opposed to an explanation for its being reasonable. See supra note 35. Dressler himself does not adopt this strategy for explaining the reasonable loss of self-control requirement. His preferred strategy for dealing with the unacceptable implications of a robust reading of the reasonable loss of self-control requirement is instead to substitute the "ordinary person" for the "reasonable person," although he believes the terms can be used interchangeably. See, e.g., Dressler, Difficult Subject, supra note 21, at 973 n.65.

83. See, e.g., Dressler, Heat of Passion, supra note 21, at 463 ("In provocation cases... the involuntariness resulting from anger is like insanity, not duress."); see also Dressler, Difficult Subject, supra note 21, at 975 n.68 ("[W]ith provocation, the killing is the result of an emotional explosion almost immediately following the provocation. The homicidal act here is the antithesis of rationality.").

84. MODEL PENAL CODE § 4.01(1) (1985). Inasmuch as the partial-incapacity theory analogizes the provoked actor to an actor who satisfies the excusing condition contained in the MPC's definition of insanity, it would seem to follow, despite the partial-incapacity theory's emphasis on the partial nature of the provoked actor's incapacity, that provocation should be a complete defense, just as insanity so defined is a complete defense.
so gripped. If so, then it doesn’t matter whether the provocation causing the desire is adequate or inadequate. Indeed, it doesn’t matter if provocation or something else entirely is its cause.

Moreover, even if the partial-incapacity theory could explain the adequate-provocation requirement, it has another problem. The theory is based on the concept of "partial incapacity." The provoked actor is entitled to mitigation because he is partially unable to control the desire to kill arising from the provocation. But the idea of a partial incapacity is incoherent. With respect to any particular task at any particular point in time an actor either does or does not have the capacity to accomplish the task. The logic of capacity is an all-or-nothing logic. Either I can φ, or I can’t φ. It may be harder or more difficult for one actor who has the capacity to φ actually to succeed in φ-ing compared to another actor who has the same capacity. Nonetheless, both actors ex hypothesi have the capacity to φ, no matter how hard they find it to exercise that capacity at a given moment in time.

85. As with the total-incapacity theory, the partial-incapacity theory could explain the adequate-provocation requirement if it were to say that an actor is responsible for his partially incapacitating desire to kill when the desire arises from inadequate provocation, but not when it arises from adequate provocation. But again, such an explanation would commit the theory to imposing punishment on the inadequately provoked actor not for what he has done, but for who he is (for his character). See infra pp. 1711–17.

86. Dressier favored abandoning the adequate-provocation requirement in his early writings. See Dressler, Provocation, supra note 21, at 475 n.46 (“I would . . . not require evidence of wrongful conduct [i.e., adequate provocation] as a prerequisite to finding manslaughter.”). In his most recent writing, however, he says that the doctrine may but need not include a requirement of adequate provocation. See Dressler, Difficult Subject, supra note 21, at 972 (describing the heat-of-passion doctrine as “potentially” containing “a justificatory feature” reflected in the adequate-provocation requirement); id. at 984 (“[I]t is perfectly appropriate to limit the partial defense to those whose lack of self-control is the result of adequate provocation and not just any provocation.”). But if the adequate-provocation requirement is treated as an optional feature of the doctrine, the question still remains: What role does the requirement play if and when the option is exercised and adequate provocation is required in order for an actor to prevail on a plea of provocation?

87. Insofar as the doctrine of diminished capacity also rests on the idea of a partial incapacity, it too seems to me to be incoherent. See infra pp. 1738–44 (discussing diminished capacity doctrine).

88. The fact that it may be harder for an actor to exercise his capacity for self-control can help explain why any such effort on his part failed on the present occasion. It can also lend credence to an actor’s claim to have tried to exercise self-control but to have failed successfully to do so on the present occasion.

89. An actor’s capacity for self-control is probably best understood as consisting of a variety of learned skills of self-control, including what might be called “brute resistance,” which consists in forming or retaining an intention to do that which one believes one ought to do. In other words, the intentional formation of the intention to do that which one believes one ought to do ultimately alters the balance of one’s desires in favor of doing that which one believes one ought to do. See MELE, supra note 25, at 58 (“A person’s capacity for self-control is a function of the modes of skilled resistance . . . and his powers of brute resistance.”); Holton & Shute, supra note 28, at 7 (“[T]here is a distinct faculty of self-control that enables agents to do what they
2. Unfair Opportunity

According to the unfair-opportunity theory, an actor who kills in the heat of passion upon adequate provocation is entitled to a partial defense, not because the desire to kill resulting from the provocation renders him totally or partially incapable of conforming his conduct to the requirements of law, but rather because expecting such conformity would be unfair. It would be too much to demand of him under the circumstances.90

Inasmuch as an actor's capacity for self-control does indeed consist of such skills, an actor's capacity for self-control might intelligibly be described as partial or diminished in a comparative sense. One actor might have available to him at any given moment fewer of the relevant skills compared to another actor. As such, it may be harder or more difficult for the first actor successfully to exercise self-control compared to the second. The first actor therefore labors, one might say, under diminished capacities for self-control compared to the second. Still, at any particular point in time an actor either will or will not have the requisite capacity for self-control such that he either is or is not able to conform his conduct to the requirements of law, assuming he cares enough to fully deploy whatever skills of self-control he has available to him.

90. Duress has been described as an excuse best explained by the unfair-opportunity branch of the choice theory of excuse. See, e.g., Joshua Dressier, Reflections on Excusing Wrongdoers: Moral Theory, New Excuses and the Model Penal Code, 19 RUTGERS L.J. 671, 710 (1988) ("Duress is a no-fair-opportunity excuse."). Unfair-opportunity theories of provocation tend to analogize provocation to duress, much like incapacity theories tend to analogize it to insanity. But the analogy between provocation and duress ultimately breaks down, as does the analogy between provocation and insanity. When duress excuses an actor for failing to conform his conduct to the requirements of law, it does so ultimately because, as the MPC puts it, the actor has acted as would a person of "reasonable firmness" or fortitude in the same situation. MODEL PENAL CODE § 2.09(1) (1985). According to the MPC:

It is an affirmative defense that the actor engaged in the conduct charged... because he was coerced to do so by the use of, or a threat to use, unlawful force against his person or the person of another, that a person of reasonable firmness in his situation would have been unable to resist.

Id. An actor who violates the law due to duress is therefore excused insofar as he is in fact permitted to act (justified in acting) as he does, not in the eyes of the law, but in the eyes of some competing system of norms incorporated into the criminal law by reference to the "person of reasonable firmness." Consequently, while a standard case of duress constitutes an "excuse" in the eyes of the law, it constitutes a "justification" in the eyes of some other normative system to which the law gives recognition.

The claim that duress is in fact some form of justification is quite common in the literature. See, e.g., ALAN WERTHEIMER, COERCION 168 (1987) (duress as "agent-relative" justification); Larry Alexander, A Unified Excuse of Preemptive Self-Protection, 74 NOTRE DAME L. REV. 1475, 1487 (1999) (duress as "personal" justification); Claire O. Finkelstein, Duress: A Philosophical Account of the Defense in Law, 37 ARIZ. L. REV. 251, 280 (1995) (duress as "agent-relative" justification); Moore, supra note 69, at 40 (unfair-opportunity branch of choice theory of excuse: "could be called the 'failed justification' idea of excuse"); Robert F. Schopp, Justification Defenses and Just Convictions, 24 FAC. L.J. 1233, 1311 (1993) (duress as justification under "conventional social morality"); see also R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS 144 (1994) (arguing that cases of duress are cases in which the "agent has not
The unfair-opportunity theory can take one of two forms. The first, like the incapacity theories, focuses on the strength or intensity of the provoked actor's desire to kill. On this account, the provoked actor faces an unfair opportunity to conform his conduct to the requirements of law if and because a reasonable person in the grip of a desire to kill as intense as that of the provoked actor would have found it especially hard to conform, even though the actor did not lack (or partially lack) the capacity to conform.

The unfair-opportunity theory in this form can account for the passion and reasonable loss of self-control requirements. As with the incapacity theories, the heat of passion is required because it constitutes, generates, or intensifies the actor's desire to kill. Likewise, the actor is entitled to the defense if and only if his failure to exercise his capacity for self-control in the face of a desire so intense is reasonable inasmuch as a person of reasonable firmness would likewise have failed to resist so intense a desire.

Unfortunately, this first form of the unfair-opportunity theory, like the total-incapacity theory, cannot account for the partiality or adequate-provocation requirements. If a reasonable person experiencing a desire to kill as strong or intense as that of the provoked actor would have killed because he would have failed to control that desire, then the provoked actor, having acted reasonably or in the same way a reasonable person would really violated the moral obligations that we accept

Craig L. Carr, Duress and Criminal Responsibility, 10 LAW & PHIL. 161, 183 (1991) (arguing that duress provides a defense in "those extraordinary cases that involve disobedience to law in response to a moral dilemma"); Joshua Dressler, Exegesis of the Law of Duress: Justifying the Excuse and Searching for Its Proper Limits, 62 S. CAL. L. REV. 1351, 1366 (1989) ("At its core the defense of duress requires us to determine what conduct we, a society of individual members of the human race, may legitimately expect of our fellow threatened humans."); Gardner, supra note 69, at 597-98 (arguing that the "gist of an excuse," including that of duress, is that one has "lived up to" the "applicable standards of character" based on some conception of one's appropriate role); Morse, supra note 36, at 341 ("Deciding which choices are too hard, that is, which threats might cause a person of reasonable firmness to yield and to do wrong [and thus be excused on the basis of duress], is of course a normative matter." (emphasis added)). But see Peter Westen & James Mangiafico, The Criminal Defense of Duress: A Justification, Not an Excuse—And Why It Matters, 6 BUFF. CRIM. L. REV. 833, 914 (2003) (arguing that an actor who validly claims duress is justified in the eyes of the law because he has "conformed to what the law ultimately regards as minimally acceptable conduct" (emphasis added)). For a brief but effective critique of the Westen-Mangiafico thesis, see Berman, supra note 18, at 72-73.

Insofar as killing under adequate provocation is permissible neither in the eyes of the law nor in the eyes of any competing system of norms to which the law is prepared to pay deference, the analogy between valid claims of duress and valid claims of provocation breaks down. If a competing system of norms to which the law was prepared to pay deference did permit an actor to kill upon adequate provocation, then provocation would be analogous to standard cases of duress. Under those circumstances, however, provocation, like duress, should be a full excuse. Some commentators who appear to analogize provocation to duress fail to explain why the latter is, but the former is not, a full excuse. See, e.g., Macklem & Gardner, supra note 22, at 824-26 (arguing that the reasonable loss of self-control requirement should vary depending on the role the actor occupies, but failing to explain why an actor who lives up to the relevant standard should not therefore be fully excused).
have acted, should be entitled to a full excuse, not a partial one. The law should not punish people for reasonable failures. Similarly, insofar as it focuses on the strength or intensity of the desire to kill, without regard to the reasons why the actor found himself gripped with such a desire, neither can the unfair-opportunity theory in this form account for the adequate-provocation requirement. If what matters is the strength or intensity of the actor's desire to kill, why he so desires is neither here nor there.

In contrast, the second form of the unfair-opportunity theory does care about why the actor wants to kill. On this account, a provoked actor faces an unfair opportunity to conform his conduct to the requirements of law if and because a reasonable person would have gotten angry at that which made the actor angry. Moreover, that anger would have constituted, generated, or intensified a desire to kill that a reasonable person would have failed to control.

This second form of the unfair-opportunity theory can arguably explain three of the doctrine's four requirements. The heat of passion is needed because it constitutes, generates, or intensifies the provoked actor's desire to kill. The adequate-provocation requirement is needed because an actor's opportunity to conform his conduct to the law is unfair if and only if the provocation he faced was adequate. Finally, the reasonable loss of self-control requirement is needed because it identifies when an alleged provocation constitutes adequate provocation. An alleged provocation is adequate if and only if a reasonable person faced with such provocation would have become so angry as to have lost self-control.

Nonetheless, the theory has two problems. First, it leaves the partiality requirement unexplained. If the adequately provoked actor's loss of self-control really was reasonable, then he should be entitled to a full defense. Second, and more ominously, it ends up punishing some provoked actors not only for what they do, but also for who they are. But punishing people for who they are, and not only for what they do, is an illegitimate end for the criminal law of a liberal state. Let me elaborate.

According to the second form of the unfair-opportunity theory, the only difference between an actor who loses self-control and kills in the heat of passion upon adequate provocation and an actor who loses self-control and kills in the heat of passion upon inadequate provocation is the adequacy of the provocation each of them confronts. Both actors kill only because they fail to control a provocation-induced desire to kill, even though each is

91. See, e.g., sources cited supra note 35.
92. Because the second form of the unfair-opportunity theory defines adequate provocation in terms of reasonable loss of self-control (and vice versa) the theory in effect embraces a consolidation strategy for dealing with the reasonable loss of self-control requirement. See supra note 35.
assumed to be fully capable of exercising such control. Nonetheless, the adequately provoked actor's failure to exercise self-control is deemed reasonable, whereas the inadequately provoked actor's failure is deemed unreasonable. The adequately provoked actor is therefore convicted of manslaughter, whereas the inadequately provoked actor is convicted of murder.

Now, if the adequately provoked actor and the inadequately provoked actor, unlike the unprovoked actor, kill only because they lose self-control in the face of provocation, and if the adequately provoked actor's claim to mitigation ultimately rests on his claim to have killed only because he lost self-control, then why is the inadequately provoked actor, who also killed only because he lost self-control, denied the chance to make the same claim?

The answer is this: The inadequately provoked actor forfeits his claim to mitigation—a claim to which he would otherwise have been entitled—because and only because he got angry for the wrong reasons. The adequate-provocation requirement therefore functions as a forfeiture rule. An actor who kills only because he loses self-control in the face of inadequate provocation should not have gotten angry in the first place, and the price he pays for his misplaced anger is to forfeit his claim to the defense.

Such forfeiture rules are well-known, though not uncontroversial, features of the criminal law. According to these rules, an actor who would otherwise have a valid defense (justification, excuse, or failure of proof) to a crime forfeits the defense (completely or partially) if he is culpably responsible for creating the conditions giving rise to it. For example, an actor who voluntarily becomes intoxicated might forfeit a claim to the effect that he could not or did not, as a result thereof, form a mental state required for conviction. Likewise, an actor who kills in self-defense, but who started the fray, forfeits any claim to self-defense, unless he successfully renounces his initial aggression. So too, an actor who places himself in a situation in which he knows he will likely be coerced to commit a crime forfeits any claim of duress. Indeed, an actor who intentionally provokes conduct that in turn provokes him to lose self-control and kill forfeits any claim of provocation. All of these actors are in some fashion responsible

93. The details of these forfeiture rules vary from jurisdiction to jurisdiction. See Paul H. Robinson, Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine, 71 VA. L. REV. 1, 2-27 (1985) (surveying then-existing approaches to "cases where an actor is in some way responsible for bringing about the conditions of his own defense"). The differences are not important for present purposes. In place of such forfeiture rules, Robinson proposes to "maintain[] the defense for the offense conduct but impose[e] liability for conduct in causing the defense conditions." Id. at 27 (capitalization altered).
94. See, e.g., MODEL PENAL CODE § 2.08(2) (1985).
95. See, e.g., id. § 3.04(2)(b)(i).
96. See, e.g., id. § 2.09(2).
97. See, e.g., MODEL PENAL CODE § 210.3 cmt. 5(a), at 64-65 (1980) (noting that it is "implicit" in the MPC formulation of the defense that "extreme emotional disturbance will not
for getting themselves into trouble. They all have “dirty hands,” as a result of which the law denies them a defense to which they would otherwise have been entitled.

The second form of the unfair-opportunity theory treats the adequate provocation requirement as another example of such forfeiture rules. An actor who fails to exercise self-control in the face of inadequate provocation forfeits any claim of provocation because he became angry when a reasonable person would not have. But if the adequate provocation requirement is treated as a forfeiture rule, then the punishment imposed on an inadequately provoked actor for murder, above and beyond that which he would have received had he been convicted of manslaughter, is punishment imposed just because he has, so to speak, a “dirty character.”

According to the second form of the unfair-opportunity theory, an adequately provoked actor’s claim to excuse is ultimately based on the simple fact that he failed to control a provocation-induced desire to kill. But an inadequately provoked actor can make the same claim. The inadequately provoked actor, like the adequately provoked actor, honestly believes the provocation facing him constitutes a serious wrong, which in turn triggers a passion-induced or -enhanced desire to kill. If so, then the only difference between the adequately provoked actor and the inadequately provoked actor ultimately lies in their respective beliefs as to the kinds of actions that constitute instances of serious wrongdoing.

Whereas the adequately provoked actor’s beliefs are true or appropriate (according to some system of norms), those of the inadequately provoked actor are false or inappropriate (according to some system of norms). The

reduce murder to manslaughter if the actor has intentionally, knowingly, recklessly, or negligently brought about his own mental disturbance); ASHWORTH, supra note 17, § 7.4(b)(ii), at 273 (noting that the effect, if any, of self-induced provocation on an actor’s claim of provocation is a question left to the jury under section 3 of the 1957 Homicide Act); A.J. Ashworth, Self-Induced Provocation and the Homicide Act, 1973 CRIM. L. REV. 483, 484-87 (discussing common law authorities). Some state statutes expressly provide that any provocation induced by the actor is inadequate as a matter of law. See, e.g., State v. Pulsifer, 724 A.2d 1234, 1238 (Me. 1999) (describing Maine law).

98. See, e.g., Ashworth, supra note 22, at 308 ("The defence of provocation implies that the loss of self-control was caused by the provocation: if the provocation was objectively slight, this suggests that the substantial cause of the loss of control was not the provocation but rather some weakness (or wickedness) in the accused’s character." (emphasis added)); id. at 311 ("[T]he primary rationale of the objective standard is that it declares the widely-felt distinction between someone who is truly provoked to kill... and someone upon whose personality a relatively trivial provocation reacts in a disproportionate way." (emphasis added)); cf Dressler, Provocation, supra note 21, at 474-75 (describing how the adequate-provocation requirement could be explained on the basis of an analogy to the forfeiture rule governing voluntary intoxication). According to Horder, the explanation Ashworth and Dressler offer or suggest for the existence of the adequate-provocation requirement means that "[d]efendants in [cases of inadequate provocation] are properly given no mitigation of the offence, because the loss of self-control ultimately stems from something for which they are 'to blame', namely, their defective character." HORDER, supra note 16, at 123 (emphasis added).
inadequately provoked actor therefore forfeits the defense, not because he acts with any more culpability than does the adequately provoked actor, but simply because he believes culpably. Moreover, if one's beliefs are constituent elements of one's character over which one has no direct or immediate control, then the additional punishment imposed on the inadequately provoked actor, who neither accepts nor otherwise identifies with the offending belief, is punishment imposed for the content of his character, not for the character of his action.

Comparing provocation with hate crimes might help clarify the point. Hate crimes can be analyzed in two very different ways. First, hate crimes

99. One might claim an actor does have control over his beliefs and desires (and thus over his character) because he chooses those beliefs and desires. As such, one might claim an actor can be held responsible for his character on the same basis as he can be held responsible for his actions. Despite the fact that Aristotle appears to have endorsed this claim, it nonetheless seems to me implausible. We do not have the same degree of control via choice over our characters as we do our actions. See, e.g., Moore, supra note 69, at 45 (doubting as an empirical matter that "we really have much capacity to mold our characters"); see also George Sher, Ethics, Character, and Action, 15 SOC. PHIL. & POL'y 1, 9 (1998) ("It is, I think, mere wishful thinking to suppose that anyone has much control over either his characteristic outlook on the world or his basic mode of engaging with others."). For efforts to defend the Aristotelian thesis, see, for example, Jonathan Jacobs, Choosing Character: Responsibility for Virtue and Vice 1, 59 (2001), claiming that "punitive sanctioning of... those who are ethically disabled (as a result of voluntary activity)" and whose "characters are [therefore] such that sound ethical considerations are inaccessible to them" is "permissible."

100. Treating the adequate-provocation requirement as a forfeiture rule also leaves the law of homicide with no way to distinguish between actors who intentionally kill without losing self-control from those who intentionally kill only because they lose self-control in the face of inadequate provocation. See, e.g., Simons, supra note 61, at 426-27. Relying on the doctrine of premeditation to distinguish between these two kinds of actors is not completely satisfying. The idea would be that actors who intentionally kill without losing self-control would be guilty of premeditated murder, whereas those who intentionally kill only because they lose self-control in the face of inadequate provocation would be guilty of simple murder. But not all intentional killers who kill without losing self-control premeditate, unless of course premeditation can be satisfied based upon a finding of intent to kill, such that any actor who intentionally kills without losing self-control can be found guilty of premeditated murder.

The premeditation doctrine is probably better understood as an awkward device through which the law delegates to the jury the discretion to single out those intentional killers who, in the jury's judgment, intentionally kill for especially reprehensible motives freely and wholeheartedly acted upon. See, e.g., Pillsbury, supra note 22, at 110 (urging that the motives elevating murder to aggravated murder be spelled out in statute and including among them "animosity toward the victim's race, religion, ethnicity, sex or sexual orientation"); Kahan & Nussbaum, supra note 23, at 325 ("[I]t is the quality of offenders' motives (or lack thereof), and not the intensity or suddenness of them, that separates the most heinous murders from the rest.").

101. See Kenneth W. Simons, Does Punishment for "Culpable Indifference" Simply Punish for "Bad Character", 6 BUFF. CRIM. L. REV. 219, 244-45 (2002) (distinguishing these two different ways in which "[c]riminal provisions authorizing enhanced punishment for bias crimes can reach" an actor).
can be analogized to crimes of specific intent. On this view, an actor guilty of a hate crime is punished more severely than an otherwise-motivated actor because he commits the crime with a specific intent (understood as a motive or desire) to humiliate or subordinate the victim, which motive or desire is triggered because the actor believes, based solely on the victim’s group membership, that the victim is inherently evil.

A hate crime so understood does not enhance an actor’s punishment simply because he happens to reveal in action his possession of the proscribed desire and the belief triggering it. Instead, his punishment is enhanced if and because his action reveals his acceptance of or


However, according to the recent analysis of Heidi Hurd and Michael Moore, hate crimes cannot plausibly be analogized to crimes of specific intent because “to explain a defendant’s action as a product of hatred is not itself to attribute to him a desire to bring about some future state of affairs,” as is true of specific-intent crimes. Heidi M. Hurd & Michael S. Moore, Punishing Hatred and Prejudice, 56 STAN. L. REV. 1081, 1122 (2004); see also Heidi M. Hurd, Why Liberals Should Hate “Hate Crime Legislation,” 20 LAW & PHIL. 215, 219-22 (2001) (earlier work defending the same claim). If I understand this argument correctly, I find it unpersuasive. Why can’t an actor guilty of a hate crime be understood to desire to bring about a future state of affairs in which the victim is humiliated or subordinated?

Hurd and Moore also appear to argue that portraying hate crimes as crimes of specific intent misdescribes what hate crimes are all about. Committing a crime against a person because you hate him simply cannot be reduced, so the argument goes, to committing a crime against him because you want to humiliate or subordinate him. If a hate crime is a crime of hate, it must necessarily include the feeling or affectivity associated with the emotion of hate. See Hurd & Moore, supra, at 1126. Thus, they claim, a hate crime is necessarily one done, so to speak, in the “heat of hate.” It seems to me, however, that at least some proponents of hate-crime legislation would be contented if such legislation defined a hate crime so as to include hate crimes qua crimes of specific intent but to exclude hate crimes qua crimes of passion.

103. Some analysts argue that intentions can be wholly reduced to a complex of beliefs and desires, see, e.g., Robert Audi, Intending, 70 J. PHIL. 387, 395 (1973), while others argue that intentions cannot be so reduced. See, e.g., Michael Brauman, Two Faces of Intention, 93 PHIL. REV. 375, 376 (1984).

104. Insofar as the emotion involved in provocation cases is anger while that involved in hate crimes is hate, the difference between these two emotions has been characterized as follows: “Anger is aroused in response to a specific, undeserved offense,” while “hate may be characterized as involving a global negative attitude toward someone considered to possess fundamentally evil traits.” BEN-ZE’EV, supra note 34, at 380; see also ELSTER, supra note 34, at 65 (“In anger, my hostility is directed towards another’s actions and can be extinguished by getting even. . . . In hatred, my hostility is directed towards another person or a category of individuals who are seen as intrinsically and irremediably bad.”). In any event, if illegitimate or unjustified hatred is not limited to hate based on group membership, Anthony Dillof makes a fair point when he says: “[T]he problem with attempting to justify bias crime statutes on a theory of greater culpability is that to the extent that greater punishments are warranted in cases of racial animus, it is because of the ‘animus,’ not the ‘racial.'” Anthony M. Dillof, Punishing Bias: An Examination of the Theoretical Foundations of Bias Crime Statutes, 91 NW. U. L. REV. 1015, 1077 (1997).
identification with that desire, which implies his acceptance of or identification with its triggering belief. Hate crimes qua specific-intent crimes can therefore be said to punish an actor's choice to express the proscribed desire and belief in action, not simply his possession of such desire and belief. As such, they punish choice, not character. Consequently, hate crimes qua specific-intent crimes cannot fairly be indicted on the charge that they punish character.\textsuperscript{105}

Second, hate crimes can be analogized to crimes of passion.\textsuperscript{106} On this view, an actor guilty of a hate crime does not wholeheartedly choose to express his hate in action. He does not accept the desire to humiliate or subordinate associated with his hate, nor does he therefore choose to express that desire or its triggering belief in action. His desire and belief are but-for causes of his action, but they are not, so it might be put, motives for his action.\textsuperscript{107} Instead, the proscribed desire and belief simply "attend" his action.\textsuperscript{108} They constitute a desire and belief he "harbor[s]" while acting.\textsuperscript{109} Consequently, hate crimes qua crimes of passion, unlike hate crimes qua crimes of specific intent, do punish an actor for his character.

Heidi Hurd and Michael Moore argue that hate crimes qua crimes of passion are objectionable precisely because the additional punishment imposed on the hate criminal, above and beyond the punishment imposed on his non-hate counterpart, is punishment imposed on him for having a bad character, or in other words, for possessing proscribed desires and

\textsuperscript{105} The worthy-motive theory of provocation can be understood as an effort to construe the provocation doctrine as defining an unusual type of specific-intent crime. According to the worthy-motive theory, the adequately provoked actor is free to act wholeheartedly on his desire to retaliate in response to the provocateur's wrongdoing, and because the law deems this desire (or motive) a relatively worthy one, he is guilty of a lesser crime than murder. Thus, whereas the adequately provoked actor wholeheartedly acts on the basis of a relatively worthy motive for which his punishment is mitigated, the actor guilty of a hate crime qua crime of specific intent wholeheartedly acts on the basis of an especially unworthy motive for which his punishment is aggravated. Again, however, the worthy-motive theory is inadequate as an explanatory theory of provocation insofar as it cannot explain the passion and reasonable loss of self-control requirements. See supra pp. 1693-95.

\textsuperscript{106} For Hurd and Moore, this second way of understanding hate crimes is the only plausible way to understand them. According to Hurd and Moore, "to explain a defendant's action as a product of hatred is... to characterize his action as a product of a particular passion within which he was gripped at the time." Hurd & Moore, supra note 102, at 1122 (emphasis added). It should be apparent that the language Hurd and Moore use to characterize hate crimes is language typically associated with the provocation doctrine. Hurd and Moore recognize the close relationship between the provocation doctrine and hate crimes qua crimes of passion. See id. at 1120 n.110 ("Inasmuch as the provocation/passion doctrine is concerned with an emotional state within which a defendant acts, it is more like the mens rea doctrine of hate/bias crimes than is any other doctrine in the criminal law.").

\textsuperscript{107} See Simons, supra note 101, at 244-47 (arguing that provocation cases and some hate crimes cases involve "causal but nonmotivational desires").

\textsuperscript{108} Hurd & Moore, supra note 102, at 1118.

\textsuperscript{109} Id. at 1123.
beliefs. At the same time, however, Hurd and Moore find nothing objectionable with the way in which the provocation doctrine mitigates an actor’s punishment based on his possession of worthy desires and beliefs. As they see it, aggravation based on bad character is objectionable, but mitigation based on good character is not.

Yet if an inadequately provoked actor, but for the inadequacy of the provocation to which he responds, would otherwise be entitled to a manslaughter conviction, then convicting him of murder is objectionable for the same reason convicting the hate-crime killer is objectionable. In both cases the proscribed desire and belief associated with the relevant emotion attending the actor’s conduct form the basis for increasing his punishment above and beyond the punishment he would otherwise have received. In neither case does the actor wholeheartedly act on the basis of the relevant desires and beliefs. In both cases the actor is therefore punished for his bad character. But imposing punishment for bad character is something a state committed to liberal principles simply does not do.

Consider in this light Commonwealth v. Carr, one of the cases described at the outset. The defendant in the case, described as a “mountain man,” was on the Appalachian Trail in Adams County, Pennsylvania when he saw two women making love. Claiming to have become enraged at the sight, he shot the women with a .22-caliber rifle. One was killed, and the other

110. See id. at 1118-29.
111. See id. at 1120 n.110 (“Inasmuch as the provocation/passion doctrine is... an exculpatory doctrine, while the mens rea doctrine of hate/bias crimes is an inculpatory doctrine, there is ultimately little analogy between the two...”). But see Simons, supra note 101, at 246 n.61 (claiming that this effort to distinguish provocation from hate crimes “fails to explain why retributive theory can justifiably recognize desires and emotional states only insofar as they exculpate”).
112. See, e.g., Hurd & Moore, supra note 102, at 1135 (“Character theories of punishment,” which “best explain[] and justif[ies]” hate crime legislation and “motivationally oriented legislation in general... cannot enjoy support from those who conceive of themselves as working within the philosophical tradition of political liberalism.”); Hurd, supra note 102, at 229 (“[T]he punishment of vice and the cultivation of virtue... are distinctively non-liberal goals.”); Jeffrie G. Murphy, The State’s Interest in Retribution, 5 J. CONTEMP. LEGAL ISSUES 283, 297-98 (1994). As Murphy says:

If a person’s [liberal] right of moral independence serves to block punishment [directly for having an illiberal character], it is hard to see why it would not also serve to block it at a later stage when the state seeks to impose, not simply the punishment proper for the offense, but some additional punishment based on [illiberal] motive or character.

Id. But cf. Andrew E. Taslitz, Condemning the Racist Personality: Why the Critics of Hate Crime Legislation Are Wrong, 40 B.C. L. REV. 739, 742 (1999) (“Hate crimes legislation promotes a vision of virtuous citizen character in a republic, a vision that requires us to condemn [through punishment] the racist personality.”).
114. POHLMAN, supra note 2, at 3.
wounded. Carr asserted provocation as a defense, but the trial judge found him guilty of first-degree murder. His conviction was affirmed on appeal. "The sight of naked women engaged in lesbian lovemaking," the appeals court held, "is not adequate provocation." Indeed, such provocation was inadequate as matter of law.

The Carr case has become a cause célèbre among academic commentators. The court's decision is widely seen as a victory for tolerance and equality over "homophobic hatred." Yet the preceding analysis suggests a different conclusion. If indeed Carr killed only because he failed to exercise self-control, then any punishment imposed on him beyond the punishment assigned to manslaughter would reflect punishment for his holding or possessing the desires and beliefs associated with homophobic hatred, or in other words, for being homophobic. If so, then while Carr may be a victory for the liberal ends of tolerance and equality, it achieved any such victory through decidedly illiberal means. It represents the exercise of state coercion to punish a citizen partly for his character, illiberal though it may be, and not only for his deeds.

115. See Carr, 580 A.2d at 1363. These are the bare-bone facts given in the reported opinion of the Pennsylvania Superior Court on appeal. A more complete account of the facts and events leading up to the crime can be found in POHLMAN, supra note 2. Insofar as Carr's putative claim of provocation rested on his having killed while (using the language of the MPC) in a state of "extreme mental disturbance," a state resulting from the sight of the women making love combined with his abnormal "psychosexual history," his claim sounds more in diminished capacity than it does provocation. Cf. id. at 238 (describing Carr as a "social predator not fully in touch with reality" (emphasis added)); Abbe Smith, Criminal Responsibility, Social Responsibility, and Angry Young Men: Reflections of a Feminist Criminal Defense Lawyer, 21 N.Y.U. REV. L. & SOC. CHANGE 433, 483 (1994-95) (giving the victim's lawyer's description of Carr as having "some sort of mental disability that, coupled with intense homophobia, exploded into homicidal violence" (emphasis added)). However, diminished capacity was not then and is not now available as an affirmative defense under Pennsylvania law, which only recognizes the defense in its so-called mens rea variant. See Commonwealth v. Legg, 711 A.2d 430, 433 (Pa. 1998); POHLMAN, supra note 2, at 102-03; see also infra note 191 (discussing two variants of diminished capacity). In any event, if in fact Carr stalked the women and attacked them with the purpose or desire of humiliating them, which purpose or desire he wholeheartedly embraced and made no effort to resist, then far from being entitled to mitigation based on provocation, he would have been guilty of a hate crime qua crime of specific intent.

117. Id. at 1364.
118. Id.
119. Kahan & Nussbaum, supra note 23, at 313. How exactly to characterize the emotion in which Carr acted is a matter of dispute. Carr's defense would no doubt have tried to characterize it as "anger" at the "wrong" he believed the women were "committing" against him. Nussbaum has characterized it as "disgust," an emotion she argues should never qualify as a predicate for provocation. MARTHA C. NUSSBAUM, HIDING FROM HUMANITY: DISGUST, SHAME AND THE LAW 130, 133 (2004).
120. Similar questions arise in which the alleged provocation is a non-violent homosexual advance. A number of commentators have called for such advances to be treated as inadequate provocation as a matter of law. See, e.g., NUSSBAUM, supra note 119, at 134 ("Because the 'reasonable man' is not simply the average man, but a normative social ideal, we should not
If an actor willfully refuses to conform his conduct to the requirements of law, the state can legitimately punish him for his defiance. But if an actor finds himself saddled with an illiberal character, one composed of illiberal beliefs and desires, and if he should be unlucky enough to lose self-control and reveal, though not accept, his character’s true colors in an otherwise criminal act, punishing him for having such a character is an illiberal use of state power. A liberal state can punish its citizens for the crimes they chose to commit, but it cannot punish them more for the lack of virtue they happen to betray in committing them, even when the virtues at stake are liberal ones. If it does so punish its citizens, it betrays its own true—and illiberal—character.

121. According to Kahan and Nussbaum, the criminal law has no choice but to judge the “values embodied in [offenders’] emotional valuations.” Kahan & Nussbaum, supra note 23, at 359. If that means the criminal law has no choice but to judge and, if appropriate, condemn an actor’s character, and if that therefore means the criminal law has no choice but to be illiberal, then for Kahan and Nussbaum, so be it. For, as they put it, “it simply could not be otherwise.” Id. at 360; cf. Nicola Lacey, Partial Defenses to Homicide: Questions of Power and Principle in Imperfect and Less Imperfect Worlds . . . ., in RETHINKING ENGLISH HOMICIDE LAW, supra note 32, at 107, 127 (“I would accept the idea that criminal law is inevitably in the business of what might be called ‘negative perfectionism’: that is, of enforcing some conception of what is beyond the purview of a good life.” (emphasis added)). In response to the objection that an actor should not be punished for his character because and insofar as he does not, in any meaningful sense of the word, choose his character, or that the content of an actor’s character is more or less a matter of good or bad luck, Kahan and Nussbaum hold out the possibility of mercy. Although an actor can be punished for his bad character, he might nonetheless also be eligible for mercy if “some unusual impediment . . . deformed the process of character formation.” Kahan & Nussbaum, supra note 23, at 360.

This analysis suffers from at least two problems. First, it is false to claim the criminal law has no choice but to assess a person’s character and, if appropriate, punish him for it. A liberal state need not and should not punish an actor simply for possessing proscribed beliefs and desires that he happens to reveal in action. It should punish him only when he accepts those beliefs and desires and expresses such acceptance in action. Second, a grant of mercy results in the complete or partial remission of an actor’s deserved punishment. But any punishment imposed on an actor for his character is, at least in a liberal state, undeserved punishment, whatever the history of the process leading to his character’s formation. An actor who the state proposes to punish for his character is therefore in need of justice, not mercy.

122. See, e.g., Hurd & Moore, supra note 102, at 1137 (“Those who persist in thinking it appropriate for the state to criminalize certain aspects of character must admit, then, that they are not liberals. Instead, they are probably political perfectionists, who, broadly speaking, consider it appropriate for the state to use its power to perfect its citizens morally.”).
The point here is not that a liberal state cannot hold its citizens in any way responsible for the content of their characters. We are responsible for our characters, if only because we are our characters, despite the fact that we cannot control who we are in the same way or to the same degree we can control what we do. A liberal state is free to establish institutions to educate its citizens in the ways of a liberal order and inculcate the laudable virtues of liberal citizenship. Indeed, it can even criticize and censure its illiberal citizens for their illiberal characters. What it cannot legitimately do is condemn them for their illiberal characters through the hard treatment of punishment. It cannot heap more punishment on a citizen who, in the course of violating the criminal law, happens also to reveal the illiberal state of his soul. A liberal state can punish the Carrs among us for failing to exercise self-control and killing in the heat of passion, but it cannot punish them more just because they get angry when a citizen of liberal virtue would not.

In the end, none of the excuse-based theories of provocation examined above can solve passion's puzzle. All of them leave pieces of the puzzle unexplained, and at least one punishes inadequately provoked actors for the bad character they display in their actions.

C. NEITHER PARTIAL JUSTIFICATION NOR PARTIAL EXCUSE

Two recent theories of provocation can be construed as promising a solution to passion's puzzle without recourse to the traditional concepts of justification or excuse. Provocation, so they claim, is neither a traditional

123. See, e.g., Meir Dan-Cohen, Responsibility and the Boundaries of the Self, 105 Harv. L. Rev. 959, 974 (1992) ("I am responsible for my character in the sense that my character's emanations and manifestations are simply generated by me. I am their author... because I am my character."); Angela M. Smith, Identification and Responsibility, in MORAL RESPONSIBILITY AND ONTOLOGY 233, 245 (Ton van den Beld ed., 2000) (claiming that "our responsibility for our attitudes—our desires, emotions, and other intentional mental states—does not flow from our decisions, but from the fact that they reflect our (implicit or explicit) judgements [sic] about reasons").

124. Kahan and Nussbaum exaggerate when they imply that refusing to punish an actor for his bad or illiberal character would "require[] us to treat large numbers of our fellow citizens as beyond the parameters of rational judgement [sic], and, indeed, to regard the whole business of shaping one's character and emotions as a matter of happenstance, rather than as part of a rational public culture." Kahan & Nussbaum, supra note 23, at 361. A liberal state is free to shape the content of its citizens' characters in a host of ways, but not through punishment.

125. See, e.g., Robert Merrihew Adams, Involuntary Sins, 94 Phil. Rev. 3, 21 (1985) ("Exactly the same responses are [not] appropriate to involuntary as to voluntary sins. In particular, only voluntary acts and omissions are rightly punished by the state."); George Sher, Blame for Traits, 35 NoUS 146, 158 (2001) ("[I]t is one thing to say that punishment for traits [i.e., character] is intelligible, and quite another to say... that it is defensible.").

126. Although I treat them here as explanatory theories put forth as proposed solutions to passion's puzzle, these two theories are perhaps better construed as normative endorsements of more restrictive formulations of the provocation defense, or as normative proposals to reform more expansive formulations, such as that of the MPC, in favor of more restrictive ones.
justification nor a traditional excuse, but something else. I pause here to examine these theories. As it turns out, neither delivers. In my view, both theories, though billed as something new, are actually partial-justification theories in disguise.  

1. Evaluative Emotions

Dan Kahan and Martha Nussbaum have recently proposed a theory of provocation they claim is neither a partial justification nor a partial excuse. The "terms of this debate"—the debate between provocation as partial justification and provocation as partial excuse—are, they say, "inadequate." Nonetheless, in the final analysis the Kahan-Nussbaum theory looks much like a traditional partial-justification theory of the worthy-motive variety.

Kahan and Nussbaum's theory begins from the premise that our emotions have a cognitive component. We experience certain emotions

127. Both theories discussed below can also plausibly be interpreted as excuse-based theories in which the adequate-provocation requirement operates as a forfeiture rule. If so, then both theories are vulnerable to the objection that they end up punishing an actor who loses self-control in the face of inadequate provocation in part because he has an illiberal or otherwise disfavored character. See supra pp. 1707-17. Kahan and Nussbaum appear willing to accept this implication. See Kahan & Nussbaum, supra note 23, at 366 (noting that the "evaluative conception" of emotion on which their theory of provocation rests "demands of people not only that they conform their conduct to a certain standard, but also that they shape their characters, and the quality of their emotions, in accordance with prevailing norms of reasonableness").

128. See Kahan & Nussbaum, supra note 23, at 318 (claiming that the "common law formulation" of the provocation doctrine is "neither" an excuse nor a justification). Kahan and Nussbaum focus their analysis on the common-law formulation of the provocation doctrine, believing the MPC formulation should be abandoned insofar as it rests on a "mechanistic" conception of emotion. See id. at 321 ("In comparison with the common law formulation, the Model Penal Code version of voluntary manslaughter is decidedly mechanistic."). Nussbaum has since characterized provocation as a partial excuse, though without elaborating on the precise nature of its excusing element. See NUSBAUM, supra note 119, at 133. In this later work, Nussbaum therefore appears to embrace an excuse-based theory of provocation in which the adequate-provocation requirement operates as a forfeiture rule.

129. Kahan & Nussbaum, supra note 23, at 318. The terms of this debate are said to be "inadequate because they embody theories of moral assessment that assume a mechanistic rather than an evaluative conception of emotion." Id.

130. See id. at 285-89 (describing the "evaluative conception" of emotion). Nussbaum finds inspiration for her cognitive theory of the emotions in the writings of the Stoics. See id. at 290 ("The Greek Stoics... developed the evaluative conception in an especially detailed and compelling manner, presenting most of the arguments that we... give... about its structure and its explanatory power."); see also John Deigh, Nussbaum's Defense of the Stoic Theory of Emotions, 19 QUINNIPIAC L. REV. 293, 296 (2000) ("The theory [of emotion] that Nussbaum advances is a modified version of the classical theory of Stoicism."). Ronald de Sousa has described this theory as "the most parsimonious type of cognitivist theory." Ronald de Sousa, Emotion, in THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., Spring 2003), available at http://plato.stanford.edu/archives/spr2003/entries/emotion/.
because we hold certain beliefs, whether those beliefs are true or false, appropriate or inappropriate. A person gets angry, for example, if and because he believes he has been wronged. We can therefore judge the appropriateness of his anger because we can judge the appropriateness (according to some system of norms) of the belief on which his anger rests. Was the actor truly wronged or not? If he was, how gravely was he wronged? Grave wrongs fairly give rise to great anger; small wrongs fairly give rise to slight anger.

Now, the law does not permit an actor to kill someone who has wronged him, no matter how gravely he has been wronged. The law generally permits private actors to use deadly force only in self-defense or in defense of others. Nonetheless, according to Kahan and Nussbaum, the law of provocation recognizes that an actor who kills upon adequate provocation is one who kills for reasons more worthy, or at least less unworthy, than one who kills upon inadequate provocation or upon no provocation whatsoever. The provocation doctrine is designed to mitigate the liability of those actors who kill for worthy reasons. So understood, the Kahan-Nussbaum theory is a species of the worthy-motive theory. As such, we should expect it to have trouble explaining the passion and reasonable loss of self-control requirements, since the relative worthiness of an actor's motive is the same whether he kills in hot blood or cold blood, or whether he tries to exercise self-control or not.

Though Kahan and Nussbaum do not address the reasonable loss of self-control requirement, they do say their theory can explain the passion.

In her most recent writings, Nussbaum acknowledges the many "kinetic" features associated with the emotions, but nonetheless argues "that all these features are not only not incompatible with, but are actually best explained by, a version of the ancient Greek Stoic view, according to which emotions are forms of evaluative judgment that ascribe great importance to things and persons outside one's control." Martha Nussbaum, Emotions as Judgments of Value and Importance, in THINKING ABOUT FEELING: CONTEMPORARY PHILOSOPHERS ON EMOTIONS 183, 185 (Robert C. Solomon ed., 2004); see also MARTHA C. NUSSBAUM, UPHEAVALS OF THOUGHT: THE INTELLIGENCE OF EMOTIONS 57 (2001) (arguing that judgments are sufficient for emotions and nothing else is necessary, although noting "we should be open-minded and humble" about this "extremely difficult question").

131. Kahan and Nussbaum appear to endorse the use of prevailing social or community norms in order to evaluate the "appropriateness" of an actor's emotion. See Kahan & Nussbaum, supra note 29, at 310 ("[C]ontemporary authorities make the adequacy of provocation an issue of fact for the jury so that the law may assess emotions against the backdrop of community mores."). At the same time, however, they urge judges to take cases away from the jury if and because the social norms a jury's members would presumably apply would be "inappropriate" because they were derived from a "bad morality." See id. at 362-65. The norms with respect to which a judge is supposed to evaluate prevailing social norms as good or bad is, so far as I can tell, left unspecified.

132. See id. at 315 ("[A]n undesirable act can be carried out for a variety of reasons and is more or less worthy of condemnation depending on what the actor's motives for doing that act express.").

133. See supra pp. 1699-95 (explaining worthy-motive theory).
As they see it, the passion requirement allows the judge or jury to assess the worthiness or appropriateness of an actor's motives or reasons for killing. An actor who kills in the heat of passion is one whose killing expresses appropriate anger; anger is appropriate if it is based on adequate provocation; and provocation is adequate if it comports with prevailing social norms. But this analysis does not explain the passion requirement. It eliminates it. Adequate provocation gives rise to appropriate anger, which provides a worthy motive for the killing, which in turn provides the basis for reducing what would otherwise be murder to manslaughter. Everything therefore depends on the adequacy of the provocation to which the actor responds, not on passion.

Kahan and Nussbaum contrast their cognitive theory of emotion, which they call the "evaluative conception," with what they style the "mechanistic conception," according to which emotions are non-cognitive "energies that impel the person to action, without embodying ways of thinking about or perceiving objects or situations in the world." On the mechanistic view, "emotions are seen as feelings relatively devoid of . . . cognitive content." According to Kahan and Nussbaum, the doctrines of the substantive criminal law have long reflected the evaluative conception, as indeed they should, since its competitor, the mechanistic conception, is a false theory of the emotions. As such, any effort to understand the criminal law with the mechanistic conception in mind is bound to produce distortion.

134. See Kahan & Nussbaum, supra note 23, at 315 ("Because it focuses on motives and not just outcomes, the evaluative view is perfectly able to explain why the traditional doctrine confines mitigation to those who experience 'heat of passion.'"). Kahan and Nussbaum also claim their theory can explain the common law's "'cooling time' limitation." See id. at 316-18. They are nonetheless "fully prepared to concede that the evaluative account of 'cooling time' is neither perfect nor exclusive." Id. at 318.

135. See id. at 315 (claiming that without the heat of passion requirement "it would be impossible to understand the defendant's act as expressing an appropriate valuation of the good . . . that is threatened by the victim's wrongful provocation").

136. See id. at 285-89 (giving a "[b]asic [a]ccount of the evaluative conception"). For Kahan and Nussbaum, particular beliefs or judgments are (1) "constituent parts" of a corresponding emotion (and not simply a cause thereof); (2) "necessary conditions" for the existence of the emotion; and (3) "most of the time anyway," "sufficient conditions" for the existence of the emotion. See id. at 293-95 (describing "different species of the evaluative conception" (emphasis omitted)).

137. Id. at 278. The "mechanistic theory," sometimes called the "feeling theory," is most prominently associated with the work of William James and Carl G. Lange. See William James, What is an Emotion?, 19 MIND 188, 189-90 (1884) ("My thesis . . . is that bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion.").

138. See Kahan & Nussbaum, supra note 23, at 279.
Our emotions are, as Kahan and Nussbaum rightly insist, cognitive. In fact, cognitivism has been described as the modern “orthodoxy” with respect to the ontology of the emotions. But if the mechanistic conception understates the cognitive, the evaluative conception has the opposite problem. It overstates the cognitive and, more importantly, understates the affective and conative. Emotions are about beliefs, but they are also about feelings and desires. Indeed, if the cognitive theory emerged as a philosophical rebellion against the narrowness of the mechanistic conception, signs of a counter-rebellion can now be detected. The goal of

139. For some scholars the “cognitive” nature of the emotions is best understood, not so much as a kind of evaluative judgment, but rather as a kind of perception, or more precisely, as a “way[] of seeing—species of determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies.” de Sousa, supra note 130; see also RONALD DE SOUSA, THE RATIONALITY OF EMOTION 172 (1987) (“Emotions ... control[] the salience of features of perception and reasoning.”); Robert C. Roberts, What An Emotion Is: A Sketch, 97 PHIL. REV. 183, 184 (1988) (characterizing emotions as “concern-based construals”); Amélie O. Rorty, Explaining Emotions, in EXPLAINING EMOTIONS 103, 113 (Amélie Oksenberg Rorty ed., 1980) (stating that emotions have “quasi-intentional components [that] form patterns of focusing and salience”).

140. See de Sousa, supra note 130; see also John Deigh, Cognitivism in the Theory of Emotions, 104 ETHICS 824, 824 (1994) (“Cognitivism now dominates the philosophical study of emotions.”). Kahan and Nussbaum claim the evaluative conception “can explain why the emotions seem to have urgency or heat: because they concern our most important goals and projects, the most urgent transactions we have in the world.” Kahan & Nussbaum, supra note 23, at 289. But this explanation for the heat or urgency associated with emotion over-intellectualizes the phenomenon.

141. According to one writer, for example, “emotions are constituted by certain cognitions, desires, and affects, such that these elements are individually necessary and, when dynamically linked, jointly sufficient for emotion.” JUSTIN OAKLEY, MORALITY AND THE EMOTIONS 7 (1992).

142. Robert C. Solomon, Emotion, Thoughts and Feelings: What is a “Cognitive Theory” of the Emotions and Does It Neglect Affectivity?, in PHILOSOPHY AND THE EMOTIONS 1, 1 (Anthony Hatimoyis ed., 2003) (describing the cognitive theory of emotion as a “revolution” to which theories emphasizing “affect programmes” are a “counter-revolution”). Expressions of this counter-rebellion can be found in, for example, PETER GOLDIE, THE EMOTIONS: A PHILOSOPHICAL EXPLORATION 4 (2000) (“[M]y position can be seen as retaining what is right about the traditional view that intentionality [cognition] is essential to emotion, but bringing in feeling in the right place, as an ineliminable part of the intentionality of emotional experience, ... not merely as an afterthought.”); DAVID PUGMIRE, REDISCOVERING EMOTION 4 (2004) (“My claim is that the attempt to understand emotions in terms of one or another type of thought [cognition], however instructive, finally fails. It either obscures the dimension of affect or misrepresents it, notwithstanding the ingenious efforts of some writers to avoid this.”); MICHAEL STOCKER, VALUING EMOTIONS 54–55 (1996) (concluding that “emotions involve affectivity,” that such affectivity is “irreducible[e],” and that the “moral importance of emotions” depends “at least in part” on the “irreducibility of such affectivity”); de Sousa, supra note 130 (“[A] consensus may be building, according to which the reaction against ‘feeling theories’ of emotion was excessive, because it was too hastily assumed that feelings could not have intentionality.”); Peter Goldie, Emotion, Feeling, and Knowledge of the World, in THINKING ABOUT FEELING, supra note 130, at 91 (claiming that cognitivism is good “so far as it goes,” but that cognitivism “leaves out... feelings... and [makes] no mention of how profoundly and systematically our emotional feelings can mislead us—of how the emotions can distort perception and reason”); Geoffrey Madell, Emotion and Feeling I, 71 ARISTOTELIAN SOCY 147,
this counter-rebellion is not to overthrow the cognitive theory, but simply to win recognition of the fact that emotion cannot live on cognition alone. Cognition is important, but feeling and desire also have an important part to play in any adequate analysis of the emotions.\textsuperscript{149}

The dichotomy Kahan and Nussbaum draw between the evaluative conception and its mechanistic competitor is overdrawn.\textsuperscript{144} Caught in its grip, they sensibly choose the evaluative and reject the mechanistic, which is the only real choice the dichotomy allows. Unfortunately, the evaluative conception’s overemphasis on the cognitive, at the expense of the affective and conative, forces Kahan and Nussbaum to give us a theory of the heat of passion with neither heat nor passion. Consequently, when all is said and done, they are left with a partial-justification theory, one in which adequately provoked actors are convicted of manslaughter instead of murder because and only because they kill for a worthy motive. All the passion is gone.

\textsuperscript{149}(Supp. 1997) ("Feeling... cannot reduce to a structure or network of judgements [sic], but nor can it be regarded as a pure epiphenomenon, a sort of affective varnish which could be cleaned off without serious consequences.").

\textsuperscript{143}. This statement is intentionally vague. I see no need to take a position on the best or correct philosophical account of the emotions. It suffices for present purposes only to insist that an acceptable analysis of the emotions must make some room for affective and conative elements capable of distorting an actor’s judgment or of causing an actor to act against his better judgment. An excellent overview of the current state of philosophical play over the emotions, together with a "summary and guide to the recent literature," can be found in de Sousa, supra note 130.


Kahan and Nussbaum’s larger argument on behalf of an “evaluative” rather than a “mechanistic” conception of emotions... seems to me to rest on an overly simple contrast: emotions are either evaluations, and hence reasonable or unreasonable, or they are brute, nonrational forces that impede moral agency. In fact, emotions are both evaluative states and states that potentially interfere with rational control in various ways.

\textit{Id.}
2. Warranted Excuse

Another "new theory" of provocation comes from Victoria Nourse. Although Nourse ultimately says provocation "remains an excuse," she does so simply "because it depends upon the particular defendant's state of mind (his emotional reactions, rather than his act)," and not because the actor lacked either the capacity or a fair opportunity to conform his conduct to the requirements of law, which are usually thought to be the hallmarks of an excuse. Moreover, although Nourse criticizes the Kahan-Nussbaum theory, her theory and the Kahan-Nussbaum theory are, truth be told, really members of the same partial-justification family. Kahan and Nussbaum give us a worthy-motive theory. Nourse gives us a disproportionate-response theory. The differences between them boil down to sibling rivalry.

---

145. See Nourse, supra note 23, at 1389.
146. Nourse describes partial-justification theories as "traditional" theories, and partial-excuse theories as "liberal" theories. She claims her "warranted excuse" theory is "new" because it "rejects both traditional and liberal views of the defense at the same time that it seeks to unify them." Id. at 1397. Again, one can interpret Nourse's theory, like that of Kahan and Nussbaum, as a partial-excuse theory in which the adequate-provocation requirement functions as a forfeiture rule. If so, then the theory again ends up punishing an actor who kills only because he loses self-control in the face of inadequate provocation for having a bad character. Indeed, I worry that punishing character is the darker and unspoken side of the "new normativity" or "new culpability" Nourse detects and celebrates in recent criminal-law scholarship. See V.F. Nourse, Hearts and Minds: Understanding the New Culpability, 6 BUFF. CRIM. L. REV. 361 (2002); Victoria Nourse, The New Normativity: The Abuse Excuse and the Resurgence of Judgment in the Criminal Law, 50 STAN. L. REV. 1435 (1998).
147. Nourse is especially critical of "liberal" or "reform" theories of provocation, exemplified most clearly in the MPC's formulation of the defense. Nourse claims "liberal" or "reform" theories are often used in practice to secure manslaughter convictions for men who kill women with whom they were having a relationship, alleging as provocation nothing other than the fact that the woman was trying to leave the relationship. See Nourse, supra note 23, at 1333 ("In the end, reform has transformed passion from the classical adultery to the modern dazing and moving and leaving. And because of that transformation, these killings, at least in reform states, may no longer carry the law's name of murder."). Nourse attempts to support this empirical claim with an analysis of reported cases. See id. at 1342-68. However, a recent analysis of provocation cases in New York County, a jurisdiction that applies the MPC formulation, challenges the truth of this claim. See Stuart M. Kirschner et al., The Defense of Extreme Emotional Disturbance: A Qualitative Analysis of Cases in New York County, 10 PSYCHOL. PUB. POL'Y & L. 102, 131 (2004) ("Contrary to the suggestions of Nourse and Kahan and Nussbaum, it appears from our data that men who kill women who have left them are not able to plead successfully an EED defense in New York.").
148. See Nourse, supra note 23, at 1398; accord id. at 1382 ("I believe that the [provocation] defense is most definitely an excuse, although not in the traditional understanding of that term."); V.F. Nourse, Reconceptualizing Criminal Law Defenses, 151 U. PA. L. REV. 1691, 1717 (2003) (characterizing provocation in its "'reform' incarnations" as "fall[ing] in the excuse or partial-excuse camp").
149. See id. at 1399-1401.
Like Kahan and Nussbaum, Nourse bases her theory of provocation on a cognitive theory of the emotions. Like Kahan and Nussbaum, Nourse bases her theory of provocation on a cognitive theory of the emotions.\textsuperscript{150} We get angry when and because we believe we have been wronged. That belief in turn can be judged true or false, appropriate or inappropriate (according to some system of norms). For Nourse, an alleged provocation is adequate, and thus provides an adequate reason to get angry, if and only if it constitutes a wrong to which the law itself would respond with violence (in the form of incarceration).\textsuperscript{151} So, while Kahan and Nussbaum would rely on social norms to judge the adequacy of an alleged provocation, Nourse would rely on the law's own norms, at least as a first approximation.\textsuperscript{152} Thus, according to Nourse, the adequately provoked actor responds to the victim's wrongdoing "with a rage shared by the law."\textsuperscript{153} As such, the law must "protect"\textsuperscript{154} an adequately provoked actor's emotions; otherwise, the law would "contradict[]" itself,\textsuperscript{155} whereas no such contradiction is involved when the law refuses to extend mitigation in cases involving inadequate provocation.

At the same time, however, the law must "condemn, at least partially, those who take the law in their own hands";\textsuperscript{156} otherwise, it would fail to "maintain its monopoly on violence."\textsuperscript{157} The defense to which the adequately provoked actor is entitled is therefore only a partial defense. The defense might also be partial, not only because the adequately provoked actor takes the law into his own hands, but also because the violence he uses in response to the victim's wrongdoing is excessive in comparison to the violence the law

\begin{footnotesize}
\begin{enumerate}
\item[150.] See id. at 1390–91.
\item[151.] See id. at 1396 ("The law only suffers contradiction [by denying a provocation defense to an actor] when it refuses to embrace a sense of outrage which is necessary to the law's rationalization of its own use of violence.").
\item[152.] Nourse says in a footnote that her "proposal does not adopt the view that unlawfulness of the provoking behavior is a sufficient test of warranted outrage." Id. at 1396 n.381 (emphasis added). Instead:

Lawfulness is [only] a guide to those kinds of outrage the law must protect. It is not a doctrinal standard. Nor should it be applied without regard to the ultimate purpose of the inquiry—to determine whether the defendant is asking us to accept a claim that permits him to legislate emotional blame vis-à-vis his victim.

Id. (emphasis added). Nourse goes on: "My argument that coherence compels the criminal law to protect the emotions to which the State itself appeals to justify its use of violence depends upon the potential use of violence, not unlawfulness in name only, nor in the violation of widely held social norms." Id. at 1396–97 n.383. Yet once all these caveats and qualifications are entered, the exact contours of Nourse's proposed test for the adequacy of an alleged provocation become difficult to discern. Cf. Horder, supra note 42, at 137 n.63.
\item[153.] Nourse, supra note 23, at 1393.
\item[154.] Id.
\item[155.] Id.
\item[156.] Id.
\item[157.] Id.
\end{enumerate}
\end{footnotesize}
itself would use. On this reading, Nourse's theory bears a striking resemblance to the disproportionate-response theory.158 Consequently, the challenge for Nourse's theory, as for any partial-justification theory, is to come up with plausible explanations for the passion and reasonable loss of self-control requirements. Nourse explains neither. With respect to the passion requirement, Nourse claims her theory can "explain why we need both emotion and judgment to explain [the provocation] defense."159 This statement might fairly be understood as a promise to explain both the passion requirement ("emotion") and the adequate-provocation requirement ("judgment"). If so, then that promise cannot be kept, for the cognitive conception of emotion on which Nourse's theory relies reduces emotion to judgment. Emotion's passion disappears, as it does in the Kahan-Nussbaum theory.160 All that's left is judgment.

158. See supra pp. 1695-98 (explaining the disproportionate-response theory). Nourse nonetheless insists her theory is not a partial-justification theory. She does so for a variety of reasons, all of which are open to dispute. First, she says a partial-justification theory cannot take account of the "defendant's perception of the triggering act," whereas her theory does. Nourse, supra note 23, at 1395. But Finbarr McAuley, an acknowledged proponent of the partial-justification theory, appears to say just the opposite. See McAuley, Anticipating, supra note 20, at 141 ("There is therefore nothing in the theory of justification which precludes a defendant from relying on a reasonably mistaken belief that this victim provoked him."). Second, she says the "defendant's characteristics" are "irrelevant to the partial justification model," but not to hers. Nourse, supra note 23, at 1395. But again, McAuley appears to say just the opposite. See McAuley, Anticipating, supra note 20, at 146-47 ("Nothing in the foregoing analysis should be taken to mean that the accused's personal characteristics are entirely irrelevant to the issue of provocation . . . other than those affecting the accused's temperament . . . ."). Third, she seems to say that partial-justification theories embrace the so-called "cooling-off" requirement, at least insofar as they exclude cases "in which emotion builds over time," whereas her theory rejects any such requirement and thus encompasses such cases. See Nourse, supra note 23, at 1395 & n.377. But, as McAuley says, partial-justification theories reject the cooling-off requirement. See McAuley, Anticipating, supra note 20, at 156 (noting that the cooling-off requirement is a requirement of the common-law doctrine, but "submitting that [under a partial-justification theory] a defendant who can show that he killed in the face of substantial provocation should, on this ground alone, be entitled to the defence").

Moreover, according to Nourse, partial-justification theorists are forced into saying "that there is something partially 'right' about killing or that some people 'deserve' to die." Nourse, supra note 23, at 1398. This is an overstatement. The most plausible partial-justification theories claim only that the victim deserved some form of retaliation in response to his wrongdoing, not that the victim deserved to die for it. Nourse also says that her "proposal does not depend upon the theory that the victim deserves to be punished." Id. at 1395. That's true if it means the victim does not deserve to be punished by the person he provoked. It's not true if it means the victim does not deserve to be punished by the state. On Nourse's theory, a victim guilty of adequate provocation does deserve to be punished by the state, at least insofar as an alleged provocation is, according to Nourse, adequate if and only if it constitutes a wrong to which the law would respond with punishment (i.e., if and only if it constitutes a crime).

159. Nourse, supra note 23, at 1397.

160. Nourse also claims her theory "requires spontaneous emotion." Id. at 1398. But why that should be so is a mystery. Such spontaneity might be needed in order to support a defendant's claim to have killed because he lacked, as Nourse puts it, "a full or fair capacity for self-control." Id. But that answer is unavailable to Nourse, since she explicitly rejects the
With respect to the reasonable loss of self-control requirement, Nourse explicitly rejects the idea that "we partially excuse because the defendant lacks a full or fair capacity for self-control." Instead, she claims, "decisions applying [the provocation doctrine] express judgments about when defendants 'should' exercise self-control."  

One assumes an actor "should," on Nourse's account, exercise self-control whenever he faces inadequate provocation. If he does not, his failure to do so is necessarily unreasonable, and courts should deny him the defense. Conversely, so it would seem, an actor is free not to exercise self-control whenever the provocation he confronts is adequate. Of course, the law might hope an actor faced with adequate provocation will nonetheless try to control his desire to kill, but he labors under no obligation to do so. If so, then any judgment as to when an actor should exercise self-control depends entirely on the nature of the provocation he faces. Actors should always exercise self-control in the face of inadequate provocation, but they need not exercise self-control in the face of adequate provocation. Consequently, all that really matters in the end is the adequacy or inadequacy of the provocation. The reasonable loss of self-control requirement, like the passion requirement, disappears.

In the final analysis, the Kahan-Nussbaum and Nourse theories of provocation are partial-justification theories traveling incognito. As such, they cannot explain the passion or reasonable loss of self-control requirements. The puzzle remains unsolved.

III. A Solution to the Puzzle

Partial-justification theories can explain the adequate-provocation and partiality requirements, but they tend to founder on the passion and reasonable loss of self-control requirements. Partial-excuse theories have the opposite problem. They can explain the passion and reasonable loss of self-control requirements, but struggle with the adequate-provocation and partiality requirements.

The challenge is to construct a theory capable of combining all the elements of provocation into a normatively attractive whole. The theory set forth below, which purports to meet this challenge, sees provocation as neither an excuse nor a justification, as those concepts are commonly understood. It sees provocation instead as a doctrine designed to mitigate the punishment of those actors who kill, not in defiance of the law, but out of ignorance or weakness. The killer with a valid provocation defense is, so

---

161. Id. It bears noting that, while the idea of a "fair" or "unfair" opportunity makes sense, the idea of a "fair" or "unfair" capacity does not. A capacity is what it is. It cannot be "fair" or "unfair."

162. Nourse, supra note 23, at 1398 (emphasis added).
PASSION'S PUZZLE

to speak, an akratic killer. Hence the name of the theory: provocation as akrasia.

A. PROVOCATION AS AKRASIA

Provocation as akrasia is based on a distinction between two forms of criminal culpability, which represent two different ways in which an actor can breach the obligations the criminal law imposes on him.

1. Two Forms of Criminal Culpability

The criminal law creates and imposes obligations. Breaching these obligations warrants condemnation through retributive punishment. In a liberal democracy, the criminal law's obligations should have a democratic pedigree. They should represent obligations we impose on ourselves for the common good. They should also be liberal in both form and content. They should speak to us clearly and should generally concern themselves with intentional or reckless acts or omissions that cause or risk causing harm to others. A liberal state does not punish an actor merely because his action gives offense, nor does it punish him merely because he gives offense, having an illiberal or otherwise disreputable character.

An actor can breach an obligation the criminal law imposes on him in two very different ways. First, he can defy it. An actor defies the law when he knows the law tells him not to \( W \), but he wholeheartedly chooses to \( W \), despite the law's demand to the contrary. In other words, an actor defies the

163. I make no effort here to defend retribution, according to which the state is either entitled or obligated to impose punishment on a criminal wrongdoer because and to the extent such an offender deserves to be punished. My general thoughts on this broader issue can be found in Stephen P. Garvey, Punishment as Atonement, 46 UCLA L. REV. 1801 (1999).


165. Jean Hampton has argued that the criminal law's concept of mens rea (guilty mind) is best understood as defiance. See Jean Hampton, Mens Rea, 7 SOC. PHIL. & POL'Y 1, 1, 24 (1990) ("[D]efiance...constitutes legal mens rea."). Hampton's conception of mens rea in the criminal law derives from a more general analysis of the nature of immorality. See Jean Hampton, The Nature of Immorality, 7 SOC. PHIL. & POL'Y 22 (1989); see also Milo, supra note 24, at 185-218 (discussing a form of immorality analogous to Hampton's conception of defiance called "preferential wickedness"). Although I agree with Hampton that defiance is one important form of criminal culpability, I disagree with her insofar as she believes (as she appears to) that defiance is the only form of criminal culpability. See Garvey, supra note 24, at 280-83 (criticizing Hampton on this ground); see also Jeremy Horder, Criminal Culpability: The Possibility of a General Theory, 12 LAW & PHIL. 193, 195-98 (1993) (same).
law if, when he acts against its demands, he does so with full knowledge of the law and full consent of the will. An actor who defies the law is, one might say, "full of will," or more simply, willful. He displays an excessive pride or hubris in himself and contempt for the law. Because he willfully chooses to pursue his own ends in defiance of the law's obligations, the defiant actor symbolically elevates himself above the law. Consequently, when a defiant actor is punished, the immediate point of doing so is to humble him, or more precisely, to humble his defiant will, and thereby issue a public statement denying his false claim to superiority.

Second, he can violate the law without defying it. An actor violates the law without defiance when either of the conditions necessary for defiance is missing. For example, an actor who honestly believes the law permits him to φ has not defied the law against φ-ing. He has acted in ignorance of the law, not in defiance. Nonetheless, such an act is culpable if the ignorance in which it is done is itself culpable insofar as the actor could and should have realized the law prohibited his action. An offense committed in honest but unreasonable ignorance of the law is therefore a culpable act, but not a defiant one.

166. See Garvey, supra note 24, at 282.
167. See, e.g., Jean Hampton, A New Theory of Retribution, in LIABILITY AND RESPONSIBILITY 377, 400 n.26 (R.G. Frey & Christopher W. Morris eds., 1991) ("I see punishment as the delivery of a defeat to the wrongdoer, as an experience of submission, of being dominated."); Herbert Fingarette, Punishment and Suffering, 50 PROC. & ADDRESSES AM. PHIL. ASS'N 499, 510 (1977) ("Punishing is the humbling of the defiant—or at least the disrespectful—will.").
168. See, e.g., Jean Hampton, An Expressive Theory of Retribution, in RETRIBUTIVISM AND ITS CRITICS 1, 12 (Wesley Cragg ed., 1992) ("Retributive punishment is a way of denying a false message about worth, and thus a way of vindicating the worth of those who have been victims of wrongdoing.").
169. If an actor's ignorance of the law or mistake of law was not only honest but also reasonable, he should, in my view, be entitled to a full excuse, as successive generations of scholars have long argued. See, e.g., Douglas Hasak & Andrew von Hirsch, Culpability and Mistake of Law, in ACTION AND VALUE IN CRIMINAL LAW 157, 173 (Stephen Shute et al. eds., 1993) ("A defendant who is ignorant of the applicable legal rule... should ordinarily... be excused if his legal mistake was a reasonable one...; or... have his punishment mitigated if he did not know but should have known of the conduct's illegality."); Gunther Arzt, The Problem of Mistake of Law, 1986 BYU L. REV. 711, 713 ("[M]ens rea... implies knowledge of a violation of the law."); Ronald A. Cass, Ignorance of the Law: A Maxim Reexamined, 17 WM. & MARY L. REV. 671, 695 (1976) (concluding that "the doctrine [of ignorantia legis] is both ineffectual and unjust"); Laurence D. Houlgate, Ignorantia Juris: A Plea for Justice, 78 ETHICS 32, 36 (1967) ("Ignorance of the law (when 'reasonable')... defeats any case for saying that a person who offends because of said ignorance obtains an unfair advantage. Hence, justice dictates his exemption."); Paul K. Ryu & Helen Silving, Error Juris: A Comparative Study, 24 U. CHI. L. REV. 421, 468 (1957) ("To be subject to [legitimate punishment] the actor... must know the rule which he violates."); cf. Dan M. Kahan, Ignorance of the Law Is An Excuse—But Only For the Virtuous, 96 MICH. L. REV. 127, 131 (1997) (arguing, contrary to the orthodox view, that "ignorance of the law is [in fact] an excuse—but only for the virtuous").

Moreover, although reasonable ignorance of the law is no excuse under prevailing doctrine, federal courts nonetheless tend to construe, where possible, the mens rea requirements of federal criminal statutes so as to diminish the risk of convicting those who act
Likewise, an actor does not defy the law if, though knowing the law prohibits φ-ing, he chooses to φ but without full consent of the will. On the one hand, such an actor wants to pursue his own ends, ends at odds with his legal obligations. On the other hand, he knows what the law expects of him, believes he should conform to its demands, and indeed, wants so to conform. But he fails to resist the desire to pursue his own ends. Consequently, he winds up acting against his better judgment and in violation of the law, but he does so without wholeheartedly embracing the law-breaking desire that his will translates into action. He could and should have obeyed the law, but he failed to do so. Like an offense committed in honest but unreasonable ignorance of the law's demands, an offense committed due to such weakness of will is a culpable act, but again, not a defiant one.

The key to solving provocation's puzzle lies in recognizing how the doctrine's pieces work together to distinguish actors who kill in defiance of the law from those who do not. In other words, provocation separates those who kill in defiance from those who kill either in a moment of ignorance, in which the actor honestly believes the law allowed him to kill, or in a moment of weakness, in which the actor's will executes a desire to kill despite his belief that he should comply with the law and despite his countervailing (though insufficient) desire so to comply. Actors guilty of murder kill in or fail to act in reasonable ignorance of their legal obligations. See, e.g., Sharon L. Davies, The Jurisprudence of Willfulness: An Evolving Theory of Excusable Ignorance, 48 DUKE LJ. 341, 344 (1998) ("'Knowledge of illegality' has now been construed to be an element in a wide variety of [federal] statutory and regulatory criminal provisions . . . ."); John Shepard Wiley Jr., Not Guilty By Reason of Blamelessness: Culpability in Federal Criminal Interpretation, 85 VA. L. REV. 1021, 1023 (1999) (stating that "[w]hen reading statutes, the Justices today suppose that Congress does not want blameless people to be convicted of serious federal crimes[,]" thereby "amend[ing] the medieval slogan that ignorance of the law is no excuse").

170. One might object that the entire phenomenon of weakness of will is, despite its familiarity, illusory. Such skepticism about the possibility of weakness of the will generally takes two forms. First, some philosophers believe putative cases of weakness of the will are really cases of ignorance, since in their view an actor never knowingly acts contrary to his better judgment. Plato famously held this position. See PLATO, The Protagoras, in THE COLLECTED DIALOGUES OF PLATO *358(c) (W.K.C. Guthrie trans., Edith Hamilton & Huntington Cairns eds., 1961). Second, some philosophers believe putative cases of weakness of the will are really cases of coercion (i.e., cases in which an actor could not have done otherwise). See, e.g., Gary Watson, Skepticism About Weakness of Will, 86 PHIL. REV. 316, 330-33 (1977). For present purposes I assume these challenges, each of which tries to reduce the phenomenon of weakness of the will to something else, can successfully be fended off. For helpful book-length treatments of the general subject, see, for example, WILLIAM CHARLTON, WEAKNESS OF WILL (1988); ROBERT DUNN, THE POSSIBILTY OF WEAKNESS OF WILL (1987); JUSTIN GOSLING, WEAKESS OF THE WILL (1990). For a recent collection of essays, which also contains a comprehensive bibliography, see WEAKNESS OF WILL AND PRACTICAL IRRATIONALITY (Sarah Stroud & Christine Tappolet eds., 2003).
defiance of the law.\textsuperscript{171} Actors guilty of manslaughter kill in ignorance or weakness.

In the end, the culpability of an actor guilty of manslaughter consists in his failure to exercise his capacity for self-control.\textsuperscript{172} Gripped with a desire to kill his provocateur, he could and should have controlled that desire. But he failed to do so. His will therefore executes his desire to kill, translating it into action. Such failures of self-control can be either \textit{direct} or \textit{indirect}.\textsuperscript{173}

First, an actor's failure to control his desire to kill can affect his action directly. The provoked actor experiences a provocation-induced desire to kill, which his will ultimately translates into action, despite his simultaneous belief that he should instead conform to the law, and despite his having some desire so to conform. Such direct failures of self-control therefore involve cases in which the actor kills in a moment of weakness, or what might be called cases of \textit{practical} akrasia. The actor fails to control the influence of desire on his \textit{action}.

Second, an actor's failure to control his desire to kill can affect his action indirectly. The direct effect of the actor's desire to kill is not on his action, but on his beliefs. Here the provoked actor's desire to kill distorts his

\\[\text{HeinOnline -- 90 Iowa L. Rev. 1730 2004-2005}\]
thinking, causing him momentarily to believe the law actually permits him to kill. Or perhaps more plausibly, his desire to kill causes him to fail to retain or access his belief that the law forbids him from killing. The actor’s will then executes his desire to kill. No conflict exists between the actor’s desire to kill and his beliefs about what the law demands of him. Instead, the actor’s desire to kill temporarily blinds him to the law’s demands. Such indirect failures of control therefore involve cases in which the actor kills in a moment of ignorance of the law, or what might be called cases of epistemic akrasia. The actor fails to control the influence of desire on his belief.

2. Putting the Pieces Together

The akrasia theory solves provocation’s puzzle only if it provides a coherent and normatively attractive explanation for each of the doctrine’s four requirements. Here is how the theory puts the pieces together.

a. Reasonable Loss of Self-Control

The theory can easily explain why the doctrine requires loss of self-control. According to the akrasia theory, the provoked actor’s failure to control his anger and desire to kill in the face of provocation constitutes his culpability. It identifies how he has failed to conform his conduct to the law’s demands, and explains why his failure is culpable. A provoked actor’s failure to exercise self-control leads him either to believe the law permitted him to kill when he could and should have realized, through the exercise of self-control, it did not, or else to execute his desire to kill when, through the exercise of self-control, he could and should have controlled it. At the same time, the loss of self-control requirement distinguishes the provoked killer from the defiant one, who kills not because he fails to exercise self-control, but because he willfully defies the law.

A provoked actor claiming the defense must not only lose self-control. His loss of self-control must also be reasonable. Unfortunately, no coherent theory of provocation can take the reasonable loss of self-control

174. One can draw a distinction between epistemic or doxastic akrasia, strictly-speaking, and self-deception. Roughly, and not uncontroversially, epistemic akrasia involves an actor’s believing \( p \) due to his desire to believe \( p \), while at the same time consciously believing he should not believe \( p \) (whether based on epistemic considerations alone or on both epistemic and non-epistemic considerations), whereas self-deception involves an actor’s believing or sincerely avowing \( p \) due to his desire to believe \( p \), while at the same time unconsciously believing not-\( p \). See, e.g., MELE, supra note 25, at 112 (offering an analysis of epistemic akrasia); ALFRED R. MELE, SELF-DECEPTION UNMASKED 50-51 (2001) (offering sufficient conditions for entering into self-deception); Robert Audi, Self-Deception, Action, and Will, 18 ERKENNTNIS 133, 137 (1982) (offering necessary and sufficient conditions for describing an agent as being in a state of self-deception); Christopher Hookway, Epistemic Akrasia and Epistemic Virtue, in VIRTUE EPISTEMOLOGY: ESSAYS ON EPISTEMIC VIRTUE AND RESPONSIBILITY 178, 185 (Abrol Fairweather & Linda Zagzebski eds., 2001) (describing “components” of a "genuine case of full-blooded [epistemic] akrasia"). The important point here is that both phenomena involve an actor’s desires or emotions influencing what he believes.
requirement at face value. The problem is simple: if an actor's loss of self-control is reasonable, then the actor should be fully excused; conversely, if losing self-control to the point of killing is never a reasonable response to provocation, no matter how grave, then no one would be entitled to the defense.¹⁷⁵

Consequently, the reasonable loss of self-control requirement cannot mean what it says.¹⁷⁶ Instead, a provoked actor's loss of self-control can be described as "reasonable" only in the limited sense that losing self-control is an all-too-ordinary human failing. Virtuous actors never experience temptation and so never need to exercise self-control. Continent actors experience temptation and thus need to exercise self-control, but being ever-continent, they never lose the battle. Alas, ordinary folks are never always virtuous, nor even always continent. They sometimes have to battle with temptation, and sometimes they lose, which is why provocation is commonly and fairly characterized as a "concession to human frailty"¹⁷⁷ or "human weakness."¹⁷⁸

b. Partiality

The theory can also explain the partiality requirement. A provoked actor, unlike an insane actor or an actor who commits a crime under duress, is not entitled to a full defense. An insane actor could not have acted otherwise, and an actor who commits a crime under duress is permitted to do so, not according to the law's norms, but according to some other system of norms to which the law defers.¹⁷⁹ In contrast, a provoked actor could have acted otherwise, and no system of norms to which the law is willing to defer would allow him to kill, not even in the face of adequate provocation. The provoked actor is therefore culpable. But his culpability is of a different order compared to that of the defiant actor. The partiality requirement recognizes this difference.

c. Passion and Adequate Provocation

What about the passion and adequate-provocation requirements? Can the theory explain them as well? According to the akrasia theory, an actor's claim to the defense is ultimately based on the fact that he killed, not out of defiance, but because he failed to exercise self-control, with his failure leading him to kill in honest ignorance of the law, or due to weakness of will.

¹⁷⁵. See, e.g., sources cited supra note 35.

¹⁷⁶. But see Macklem & Gardner, supra note 22, at 825 n.21 (insisting that "[w]hen the standard of self-control in provocation is said to be the standard of the merely 'ordinary' person, this must be interpreted to mean the ordinarily reasonable person" but offering no explanation for why a valid plea of provocation should not then be a full excuse).

¹⁷⁷. Uniacke, supra note 19, at 13.

¹⁷⁸. Dressler, Heat of Passion, supra note 21, at 442.

¹⁷⁹. See sources cited supra note 90.
As such, the passion and adequate-provocation requirements might appear superfluous. All that really matters, so it might appear, is the fact that the actor killed only because he failed to exercise self-control.

But the passion and adequate-provocation requirements do have a role. The akrasia theory portrays provocation as a doctrine designed to identify actors who kill only because they fail to exercise self-control. The passion and adequate-provocation requirements enter this picture, not as substantive rules, but rather as evidentiary rules. An actor who claims to have tried but failed to exercise self-control is, all else being equal, simply more believable if he kills upon adequate provocation in the heat of passion. The passion and adequate-provocation requirements therefore corroborate the provoked actor's underlying claim of loss of self-control upon which the defense of provocation is ultimately based.

180. See, e.g., 2 SIR GERALD H. GORDON, THE CRIMINAL LAW OF SCOTLAND ¶ 25.34, at 354 (Michael G.A. Christie ed., 2001) (“Instead of being used as a way of testing the truth of the accused’s statement that he lost self-control, the reasonable man has been turned into an objective standard of self-control.”); HART, supra note 28, at 35 (“Further difficulties of proof may cause a legal system to limit its inquiry into the agent’s ‘subjective condition’ by asking . . . ‘whether a reasonable man in the circumstances would have been deprived (say, by provocation) of self-control . . . .’”); J.W. CECIL TURNER, KENNY'S OUTLINES OF CRIMINAL LAW § 121, at 176 (19th ed. 1966) (arguing that the common-law categories of adequate provocation "began as a matter of evidence[, but later] crystallized into a rule of law and became an objective legal test of liability"); 1 J.W. CECIL TURNER, RUSSELL ON CRIME 534 (12th ed. 1964) (stating that the use of the reasonable man in the law of provocation is “one more illustration of the way in which a point of evidence has been allowed to slide into a point of law, and of the inevitable mischief which thereby results”); Dennis Klimchuk, Outrage, Self-Control and Culpability, 44 U. TORONTO L.J. 441, 454-55 (1994) (explaining how the adequate-provocation requirement can be “employed . . . to determine whether the accused’s claim to have lost self-control was authentic”); Weber, supra note 21, at 159 (“Possibly the requirement of [adequate] provocation can be justified on the basis that it is necessary to insure that the heat of passion is genuine . . . .”); Herbert Wechsler & Jerome Michael, A Rationale of the Law of Homicide I, 87 COLUM. L. REV. 701, 717-18 (1937) (“The English courts . . . regarded only a few acts . . . as provocative, apparently on the ground that they alone resulted in emotional pressure under which ordinarily law abiding men might be expected to collapse.”).

Kahan and Nussbaum find this portrayal of the adequate-provocation requirement “implausible,” because the common-law categories, “if they were merely generalizations about the kinds of offenses that typically destroy volition, . . . seem manifestly underinclusive.” Kahan & Nussbaum, supra note 23, at 308. As Kahan and Nussbaum observe, “[I]t seems implausible, for example, to think that nineteenth-century decisionmakers would have viewed rage as a surprising or abnormal reaction to the infidelity of a man’s fiancée.” Id. But insofar as the decisionmakers to which Kahan and Nussbaum refer viewed provocation as a partial justification, they would naturally not have seen the common-law categories of adequate provocation as “merely generalizations about the kinds of offenses that typically destroy volition.” Id. On the contrary, those categories would have represented legally recognized reasons for one citizen intentionally to kill another falling short of self-defense or defense of others.

181. This interpretation of the adequate-provocation requirement has a respectable pedigree. According to Horder's historical analysis of English common law, the common-law categories of adequate provocation, though originally “important as illustrations of the element of partial justification, of cases where the defendant has not gone too far beyond the mean,”
Under the akrasia theory, an alleged provocation is adequate if applicable social norms would permit an actor to use non-lethal violence in response to it, or at the least would permit some form of overt response, which in the heat of passion could understandably spill over into lethal violence. The law does not permit the use of any such violence, lethal or non-lethal. Nor can the applicable social norms be understood to permit the use of lethal violence. If they did, then provocation would be a full defense akin to duress. The problem arises when the adequately provoked actor’s desire to respond to the provocation in a socially sanctioned manner combines with the heat of passion. Under these circumstances the result may well be a new and intense desire: a desire to use lethal violence, which neither the law nor prevailing social norms condone. The provoked actor can and should control this desire, but he fails to do so. He therefore ends up executing his provocation-induced desire to kill.

---

were interpreted by the common law courts from roughly the mid-eighteenth to the close of the nineteenth century as “important, instead, for no more than evidentiary reasons.” HORDER, supra note 16, at 89.

182. Relying on social norms naturally raises a number of difficult questions. How should those norms be identified? In a multi-cultural society, whose norms count when the norms of different groups conflict? What if those norms are changing? What if those norms are morally retrograde? Questions such as these tend especially to come to the fore in so-called cultural-defense cases. See, e.g., ALISON DUNDES RENTELN, THE CULTURAL DEFENSE 31–36 (2004) (discussing the use of cultural evidence in provocation cases). The akrasia theory can provide no easy answers to these questions, but it does help identify the basic question one should be asking in such cases: Did the actor kill out of honest ignorance of the law or weakness of will, or did he kill out of defiance?

183. For efforts to explain how emotions can produce akratic action, see, for example, DE SOUSA, supra note 139, at 200 (stating that insofar as the “essential role [of the emotions] lies in establishing specific patterns of salience[,]... they are perfectly tailored for the role of arbitrators among” competing beliefs and desires and thereby explaining weakness of the will); ELSTER, supra note 34, at 350 (noting that “[e]motions may affect behavior in the present by virtue of the associated arousal (viscerality) that makes one act against one’s own better judgment,” or they “may affect beliefs and, through them, behavior”); GOLDIE, supra note 142, at 111 (stating that emotions can form “the basis for explaining certain sorts of weakness of the will or akrasia”); Christine Tappolet, Emotions and the Intelligibility of Akratic Action, in WEAKNESS OF WILL AND PRACTICAL IRRATIONALITY, supra note 170, at 97, 98 (“[S]ince emotions can be seen as perceptions of values... they have the capacity to make akratic action intelligible, as distinct from merely causally explainable.”). Inasmuch as emotions increase the salience of one course of action over another, Amelie Rorty’s observation that the “akratic alternative sometimes acquires its attractions by its power to dominate the agent’s attention” is also pertinent. Amelie Oksenberg Rorty, Akrasia and Conflict, 23 INQUIRY 193, 209 (1980).

An actor who pleads provocation is asking the finder of fact to believe him when he claims to have tried but failed to exercise control over his desire to kill, such that when he killed he did so in a moment of ignorance or weakness. Either way, his claim is more believable if the provocation facing him was adequate and if he killed in the heat of passion. Conversely, an actor who claims to have tried but failed to control his desire to kill is less believable if he kills in the face of inadequate provocation or without being in the heat of passion.1

Treating the adequate provocation requirement as an evidentiary rule means not treating it as a forfeiture rule.184 Consequently, the akrasia theory is not forced to confess, when it denies the defense to the inadequately provoked actor, that it does so because he has a malformed character. The akrasia theory denies the defense to the inadequately provoked actor, not because his character is reprehensible, but because his claim to have lost self-control is incredible. An actor who kills in the face of inadequate provocation is therefore punished for murder, not because his character is malformed, but because the law assumes he killed in defiance of its demands, any claim he might make to the contrary notwithstanding.

Assigning the adequate-provocation requirement an evidentiary function, and not a forfeiture function, saves the provocation doctrine from punishing the inadequately provoked actor for having a bad character. But it has its own costs. As an evidentiary rule, the adequate-provocation requirement will deny the defense to an actor who, though he kills only because he loses self-control, does so in the face of inadequate provocation. In principle, however, such an actor should be entitled to the defense. Consequently, the adequate-provocation requirement renders the provocation doctrine under-inclusive. Thanks to the adequate-provocation requirement cum evidentiary requirement, the defense will sometimes be withheld when it should be extended.

The adequate-provocation requirement's wisdom depends on the balance between its costs and benefits.185 On the cost side, application of the rule will inevitably produce some false positives. It will convict for murder

---

184. A provoked actor is also less believable if he experiences no remorse when his passion subsides. When a provoked actor regains his senses, he should look aghast upon what he has done. He should feel ashamed for having let his desire to kill distort his understanding of the law's demands, or for not having gathered the fortitude he could and should have gathered to comply with those demands. See, e.g., G.R. Sullivan, Anger and Excuse: Reassessing Provocation, OXFORD J. LEGAL STUD. 421, 425 (1993) ("[T]he essence of the defence is revealed in rash conduct, conduct which the ethically well-disposed agent will subsequently regret but for which sympathy, even on occasion empathy, may be attracted.").

185. See supra pp. 1709-17.

186. See FREDERICK SCHAUER, PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE 152 (1991) ("[T]he design of any decision-making procedure involves an assessment of the comparative frequency and consequences of the different errors.").
those actors who, though they kill in the face of inadequate provocation, do so only because they lose self-control, and who should therefore only be convicted of manslaughter. On the benefit side, application of the rule will preclude some false negatives. Without an adequate-provocation requirement, an actor who kills in the face of inadequate provocation without losing self-control may nonetheless be able to persuade a jury otherwise, thus securing an unwarranted conviction for manslaughter. If the false negatives precluded offset the false positives produced, then the adequate-provocation requirement has something to be said for it. Otherwise, considerably less can be said.\footnote{187}

One final issue remains. According to the akrasia theory, an actor who kills only because he fails to exercise self-control should be treated differently from one who kills in defiance. The provocation defense translates this general principle into legal doctrine. But now consider how the doctrine should apply when the actor pleading provocation is fairly described as "hot-headed," "bad-tempered," "exceptionally pugnacious," or "irascible." Such an actor characteristically or routinely fails to exercise self-control. As such, even the most defendant-friendly approaches to

\footnote{187. Eliminating the adequate-provocation requirement would be a step in the direction of transforming the provocation doctrine into a generic partial defense. See Simester & Sullivan, supra note 17, at 360 ("There is much to be said for the abolition of the defence of provocation and its replacement by more general and straightforward grounds of extenuation."). Stephen Morse has recently proposed one version of such a generic partial defense, which he categorizes as a partial excuse, based on an actor's diminished capacity for "rationality" or "normative competence," both of which are, according to Morse, "continuum concepts." See Morse, supra note 36, at 340, 397; see also Jeremy Horder, Pleading Involuntary Lack of Capacity, 52 CAMBRIDGE LJ. 298, 316 (1993) (urging the creation of a "diminished responsibility" verdict that would be available as a general defence"); R.D. Mackay & B.J. Mitchell, Provoking Diminished Responsibility: Two Pleas Merging into One?, 2003 CRIM. L. REV. 745, 757–58 (proposing an MPC-like manslaughter provision designed to merge provocation and diminished capacity into a single plea according to which an actor's "extreme emotional disturbance" and/or "unsoundness of mind" are such that "the offence ought to be reduced to one of manslaughter"); cf. Alexander, supra note 90, at 1494 (proposing a generic excuse, as contrasted with a generic partial excuse, described as a "unified excuse of preemptive self-protection" modeled on the MPC's duress provision). Morse also offers various reasons why "[c]onsidering partial responsibility only at sentencing . . . suffers from defects." Morse, supra note 36, at 398.

Morse originally proposed modeling the language of his proposed generic partial defense along the lines of the MPC's "extreme mental or emotional disturbance" manslaughter provision. See Morse, supra note 36, at 400. But cf. Dressler, Difficult Subject, supra note 21, at 986 n.114. ("In my view, any generic partial excuse ought to be more explicit than the EMED language regarding the 'central excusing notion.'"). Morse has more recently modified the language of his proposal. See Morse, supra note 32, at 300 (proposing alternative formulation according to which "[t]he jury may find the defendant [guilty but partially responsible] if, at the time of the crime, the defendant suffered from substantially diminished rationality for which the defendant was not responsible and which substantially affected the defendant's criminal conduct"). Morse notes that his proposal has been "substantially influenced" by a similar proposal advanced several years ago by Herbert Fingarette and Anne Fingarette Hasse. See id. at 299 n.23 (citing Herbert Fingarette & Ann Fingarette Hasse, Mental Disabilities and Criminal Responsibility 254–57 (1979)).}
provocation hesitate to extend him the benefit of the defense. Nonetheless, insofar as the hot-headed actor kills only because he fails to exercise self-control, isn’t the akrasia theory forced to afford him the defense?

Not necessarily. The akrasia theory requires not only that the actor have the capacity to control his desire to kill. It also demands that the actor try to exercise that capacity. If, as might reasonably be supposed, the so-called hot-headed actor is one who characteristically or routinely fails to exercise self-control because he does not even try to exercise it, or just does not care enough to try to exercise it, then he has little, if any, claim to our sympathies, nor should he have any claim to the defense. On the contrary, the hot-headed actor, so understood, can be fairly grouped with the defiant actor. The moral difference between an actor who does not care enough to

188. See, e.g., Regina v. Smith, [2000] 4 All E.R. 289, 318 (H.L.). The court stated:

Society should require that [an actor] exercise a reasonable control over himself, but the limits within which control is reasonably to be demanded must take account of characteristics peculiar to him which reduce the extent to which he is capable of controlling himself. Such characteristics as an exceptional pugnacity or excitability will not suffice.

Id. (opinion of Lord Clyde).

189. The distinction between an actor who tries but fails to exercise self-control and one who does not even try raises a “certain paradox about self-control.” See MELE, supra note 25, at 62. An actor will need to exercise self-control when he believes he should φ but lacks sufficient motivation or desire to φ at the moment of acting. But if he lacks sufficient motivation or desire to φ, and instead wants to not-φ more than he wants to φ, then from where does he get his desire or motivation to exercise his capacity for self-control so as to alter his balance of motivation in favor of φ-ing? Philosophers have offered various solutions to this paradox. See, e.g., id., at 55 (resolving the paradox by reference to “compound” first-order desires); JEANETTE KENNEDY, AGENCY AND RESPONSIBILITY: A COMMON-SENSE MORAL PSYCHOLOGY 139 (2001) (resolving the paradox by reference to “cognitive habits”); John Bigelow et al., Temptation and the Will, 27 AM. PHIL. Q. 39, 44 (1990) (resolving the paradox by reference to second-order desires); Richard Holton, How is Strength of Will Possible?, in WEAKNESS OF WILL AND PRACTICAL IRRATIONALITY, supra note 170, at 39, 58 (resolving the paradox by reference to a distinct “faculty of will-power,” which is “something like a muscle”); Jeanette Kennedy & Michael Smith, Philosophy and Commonsense: The Case of Weakness of Will, in PHILOSOPHY IN MIND 141, 151 (Michaelis Michael & John O’Leary-Hawthorne eds., 1993) (resolving the paradox by reference to “cognitive dispositions”); Jeanette Kennedy & Michael Smith, Frog and Toad Lose Control, 56 ANALYSIS 63, 70 (1996) (same); R. Jay Wallace, Addiction as Defect of the Will, 18 LAW & PHIL. 621, 636 (1999) (resolving the paradox by reference to “volition,” i.e., “a kind of motivating state that...[is] directly under the control of the agent”); R. Jay Wallace, Three Conceptions of Rational Agency, 2 ETHICAL THEORY & MORAL PRAC. 217, 236 (1999) (same). For present purposes it seems unnecessary to enter this debate.

190. See, e.g., Ashworth, supra note 22, at 309–10 (implying that the “bad-tempered man” described in English caselaw and commentary is one who fail[s] to make [an] effort [to control his temper]” (emphasis added)). Describing an actor as “hot-headed,” “bad-tempered,” or “irascible” can of course carry other meanings. See, e.g., HORDER, supra note 16, at 103 (“[B]ad-tempered people consistently go wrong in their judgments of wrongdoing by far too hastily judging that they have been wronged, and by judging that they have been wronged much more seriously than they really have,”).
try to control his desire to kill (the hot-headed actor) and one who embraces it (the defiant actor) seems pretty slim. Or to put it another way, an actor who, due to indifference, exercises no effort to control a desire to kill arguably belongs in the same moral category as one who embraces or accepts that desire.

B. DIMINISHED CAPACITY

Provocation is not the only legal doctrine designed to mitigate murder to manslaughter. The doctrine of diminished capacity, also known as diminished responsibility or partial responsibility, performs the same job. No effort to solve passion's puzzle would be complete without at least some comment on how these kindred doctrines relate to one another. Examining the relationship between provocation and diminished capacity should also shed additional light on the akrasia theory.

The surface differences between the two doctrines are twofold. First, provocation requires something to have provoked the actor. Diminished capacity requires no such provocation. Second, whereas provocation is a doctrine upon which any psychologically normal or ordinary actor can rely,

191. Two distinct doctrines travel under the label "diminished capacity," usually referred to in the American literature as the "mens rea variant" and the "partial responsibility variant." The mens rea variant permits the use of evidence of mental abnormality to negate a mental state required for conviction, i.e., to establish a failure-of-proof defense. The partial-responsibility variant permits the use of evidence of mental abnormality to establish a partial excuse. See, e.g., Stephen J. Morse, Diminished Capacity, in ACTION AND VALUE IN CRIMINAL LAW, supra note 169, at 299, 240-42 [hereinafter Morse, Diminished Capacity] (explaining the distinction); Stephen J. Morse, Undiminished Confusion in Diminished Capacity, 75 J. CRIM. L. & CRIMINOLOGY 1, 1 (1984) (same). The doctrine discussed in the text is the partial-responsibility variant. Although I use the term diminished capacity in deference to what appears to be the prevailing convention in the American literature, the concept of diminished capacity suffers in my view from the same conceptual incoherence as does the idea of partial incapacity. See supra pp. 1704-05.

192. See 1 ROBINSON, supra note 17, § 101, at 474. The doctrine in the U.K. is known as "diminished responsibility," codified as section 2 of the English Homicide Act of 1957. See Homicide Act, 1957, c. 11, § 2 (Eng.) ("Where a person kills... another, he shall not be convicted of murder if he was suffering from such abnormality of mind... as substantially impaired his mental responsibility for his acts and omissions in doing... the killing."). For helpful overviews of the history and application of the English doctrine, see R.D. MACKAY, MENTAL CONDITION DEFENCES IN THE CRIMINAL LAW 180-206 (1995); EDWARD E. TENNANT, THE FUTURE OF THE DIMINISHED RESPONSIBILITY DEFENCE TO MURDER 1-70 (2001); R.D. Mackay, Diminished Responsibility and Mentally Disordered Killers, in RETHINKING ENGLISH HOMICIDE LAW, supra note 32, at 55.

193. Although recognized in section 2 of the English Homicide Act of 1957, "[f]ew [American] states" recognize the partial-responsibility variant of diminished capacity. DRESSLER, supra note 17, § 26.03, at 366. Nor is it recognized under federal law. See, e.g., United States v. Pohlot, 827 F.2d 889, 890 (3d Cir. 1987) ("We conclude that... Congress intended... to prohibit the defenses of diminished responsibility and diminished capacity."). The MPC's manslaughter provision was intentionally crafted to encompass both provocation and diminished capacity. See MODEL PENAL CODE § 210.3 cmt. 5(a), at 54 (1980) (stating that the language of the MPC's manslaughter provision "may allow an inquiry into areas which have been treated as part of the law of diminished responsibility or the insanity defense").
diminished capacity is reserved for actors who are in some way psychologically abnormal. Consequently, while evidence of the diminished actor’s psychological profile is relevant to his plea for mitigation, evidence related to the provoked actor’s psychological profile is not. Unlike his diminished counterpart, the provoked actor is supposed to be as normal as you and me.

But the differences between the two doctrines are only at the surface, since the ground upon which a provoked actor and a diminished actor claim mitigation should be seen as common ground. Each admits he violated the law; moreover, each claims to have done so only because at the moment of acting he either made an honest mistake of law or else failed due to weakness of will to follow his better judgment in favor of compliance. Consequently, the provoked actor and the diminished actor differ not so much in the substance of their plea as in the story each tells and in the evidence

194. See, e.g., Dressler, Difficult Subject, supra note 21, at 985 (“Provocation deals with the emotions and actions of ordinary persons, whereas diminished capacity relates to the thinking processes and actions of unordinary persons. Provocation deals with ordinary human weaknesses, while diminished capacity focuses on special weaknesses, on illnesses and pathologies.”); Dressler, Heat of Passion, supra note 21, at 459 (“Diminished capacity involves a mental disturbance which peculiarly involves the killer. Heat of passion is a concession to human weakness, to a universal human condition.”). As used in section 2 of the 1957 English Homicide Act, the phrase “abnormality of mind” has been read wide enough to cover the mind’s activities in all its aspects, not only the perception of physical acts and matters, and the ability to form a rational judgment as to whether an act is right or wrong, but also the ability to exercise will power to control physical acts in accordance with that rational judgment.


each musters to support and substantiate it. The differences between the two pleas boil down to proof and procedure.

The provoked actor tells everyman's story. He has no need to emphasize the details of his own psychology in order to buttress his claim to have tried, albeit unsuccessfully, to control his desire to kill. According to the provocation story, anyone faced with the kind of provocation the provoked actor faced might very well have failed in the heat of the moment successfully to have exercised his powers of self-control, such that his emotions and desires ended up getting the better of him.

In contrast, the diminished actor tells a very personal story. He uses mental health experts to describe the minutiae of his mental and emotional makeup. The point of introducing such evidence is not to show that he could not know or conform his conduct to the law's demands. If his problems reached those depths, then he would be insane. The point is instead to persuade the trier of fact that his various mental abnormalities or deficiencies made it especially difficult, though not impossible, for him to understand the law, or if he did understand it, to conform his conduct accordingly. Compared to the provoked actor, the diminished actor needs to reveal much more about his personal biography in order to persuade a fact-finder to believe his claim to have killed in ignorance or weakness, and not in defiance.  

Some commentators nonetheless insist or imply that an important moral distinction exists between provocation and diminished capacity. Diminished capacity, so the argument goes, is a plea no self-respecting moral agent would enter. As a so-called "partial exemption" from criminal

196. The diminished-capacity story and the provocation story can of course be combined in an appropriate case. An actor may point to some provocation on the part of the victim and to some mental infirmity or abnormality making self-control harder for him than for those without such an infirmity or abnormality. Because provocation and diminished capacity share the same basis, any such story is in my view a perfectly coherent one to tell. See, e.g., Mackay, supra note 192, at 74 ("[D]espite the fact that a verdict on the basis of both [provocation and diminished responsibility] has been described as 'surely illogical' it has become clear that the two pleas are not mutually exclusive.").

197. See, e.g., Dressler, Heat of Passion, supra note 21, at 460 ("If the criminal law is to coherently deal with the provocation defense, any attempt to define or understand heat of passion should be conducted separately from diminished capacity."); Gardner & Macklem, Compassion, supra note 195, at 627 ("Plea like diminished responsibility and insanity... are reserved[, unlike a plea of provocation,] for those who are not quite among us, who cannot provide an intelligible account of themselves, and whose susceptibility to the full range of human judgment is therefore in doubt."); Jeremy Horder, Between Provocation and Diminished Responsibility, 10 KINGS C. L.J. 143, 147 (1999) (arguing that the doctrines of provocation and diminished capacity should be construed so as to "preserve the moral integrity of each defense"); Macklem & Gardner, supra note 22, at 827 ("[T]he whole point of pleading provocation rather than diminished responsibility is to garner the respect and self-respect that flows from being judged by the proper standards.").
liability, it amounts to a confession of inadequacy and abnormality. The diminished actor is an incomplete moral agent, and as such an incomplete person. Provocation, in contrast, carries no such stigma. Any self-respecting actor might, through no fault of his own, find himself having killed in the

198. Those who characterize diminished capacity as a partial exemption usually characterize insanity and infancy as complete exemptions. Others characterize insanity and infancy as status defenses. See Michael S. Moore, Causation and the Excuses, 73 CAL. L. REV. 1091, 1098 (1985) (arguing that infancy and insanity are both status defenses). Consequently, one might fairly suppose that exemptions are indistinguishable from status defenses. But see John Gardner, The Mark of Responsibility, 23 OXFORD J. LEGAL STUD. 157, 162 (2003) (arguing that while infancy is a status defense insanity is an exemption and not a status defense).

Although some commentators apparently see no real harm in treating insanity as a status defense or exemption (assuming they mean the same thing), see, e.g., Duff, supra note 78, at 62 n.47 (suggesting that not much turns on how insanity is classified so long as one recognizes that insanity "is indeed a very different kind of excuse from such excuses as duress," since neither "share[s] just the same logic and structure"); Morse, Diminished Capacity, supra note 191, at 247 n.38 (noting that "not . . . much turns on" distinguishing "between cognitive and volitional excuses, such as mistake and duress on the one hand, and status excuses, such as infancy and insanity, on the other"), others argue that the status-exemption theory of insanity is over-inclusive. See, e.g., Anthony Kenny, Can Responsibility be Diminished?, in LIABILITY AND RESPONSIBILITY 13, 24-25 (R.G. Frey & Christopher W. Morris eds., 1991) ("Treating madness as a status rather than a factor . . . gives a certified mental patient a license which is not given to others: he knows that there are certain things which he may do without being held criminally responsible, while all others not of the same status will be held responsible."); Christopher Slobogin, A Rational Approach to Responsibility, 83 MICH. L. REV. 820, 832 (1985) (book review) ("[S]urely . . . a person can murder for irrational reasons one day and rob for rational reasons the next.")

In contrast to the status-exemption theory of insanity, the traditional theory of insanity calls an actor insane if, as a result of mental disease or defect, he lacked the capacity to know or conform his conduct to the requirements of law. In other words, he acts while in an excusing condition (lack of capacity) for which he is presumably not culpably responsible, arising as it does from a mental disease or defect. If the status-exemption theory is over-inclusive, then the traditional incapacity theory is arguably under-inclusive, inasmuch as it fails to account for all cases in which we would intuitively describe the actor as "insane." See, e.g., Michael S. Moore, Mental Illness and Responsibility, 39 BULL. MENNINGER CLINIC 308, 317 (1975) (making this claim against the traditional incapacity theory).

One sensible response to the under-inclusiveness of the traditional incapacity theory and the over-inclusiveness of the status-exemption theory is to say an actor is insane if the particular beliefs or desires on which he acts in committing a crime can be characterized as irrational. See, e.g., Joel Feinberg, What Is So Special About Mental Illness?, in DOING AND DESERVING 272, 285 (1970) ("We get closer to the heart of the matter, I think, if we say that the mentally ill criminal's motives are unintelligible because they are irrational."); Slobogin, supra, at 831 ("For me, the unintelligibility of reasons [beliefs and desires] is a sufficient basis for insanity."). But see Christopher Slobogin, An End to Insanity: Recasting the Role of Mental Disability in Criminal Cases, 86 VA. L. REV. 1199, 1200 (2000) (arguing that "insanity should be eliminated as a separate defense, but that . . . mental illness should be relevant in assessing culpability only as warranted by general criminal-law doctrines concerning mens rea, self-defense, and duress"). This theory is narrower than the status-exemption theory (since it looks to the irrationality of the actor's beliefs and desires, not to the irrationality of the actor as such) and broader than the traditional incapacity theory (since it looks to the irrational content or nature of the actor's beliefs and desires, not simply to the actor's capacity to know the law or conform to its requirements).
heat of passion upon adequate provocation. In short, the argument concludes, diminished capacity, unlike provocation, is a shameful plea to make.\textsuperscript{199}

Yet unless a diminished actor's mental or emotional abnormalities are themselves features about which the actor should be ashamed, diminished capacity should not be seen as a shameful plea. Again, the diminished actor, unlike the insane actor, is not claiming that mental disease or defect rendered him incapable of knowing or obeying the law. Nor is he claiming to be \textit{exempt} from blame or punishment because he fails to qualify as a moral agent. Nor is he even claiming to be \textit{partially exempt}. Indeed, the concept of a partial exemption is incoherent.\textsuperscript{200} An actor is either eligible to participate in the criminal law's game of praise and blame, or not. It makes no sense to say an actor is half-in and half-out.

Instead, the diminished actor is claiming his mental and emotional makeup made it harder, but not impossible, for him to know and conform his conduct to the law's demands. The trier of fact should therefore believe him when he says that he violated the law out of momentary ignorance or weakness. Moreover, unless the diminished actor is culpably responsible for the mental or emotional abnormalities under which he labors, he has no reason to be ashamed of them.\textsuperscript{201} The only thing about which he should be ashamed is his failure to control his desire to kill, a desire he could and should have controlled in spite of his mental and emotional baggage. But that failure is one he shares with the provoked actor.

Because neither the provoked nor the diminished actor defies the law, the point of punishing them cannot be to humble a defiant will.\textsuperscript{202} When the

\textsuperscript{199} John Gardner and Timothy Macklem are the most forceful defenders of this position. \textit{See}, \textit{e.g.}, Gardner \& Macklem, \textit{Compassion}, supra note 195, at 627 ("Pleas like diminished responsibility . . . are reserved for those who are not quite among us . . . . [S]elf-respecting people would rather that their cases fell under provocation than under diminished responsibility."); Macklem \& Gardner, \textit{supra} note 22, at 827 ("[T]he whole point of pleading provocation rather than diminished responsibility is to garner the respect and self-respect that flows from being judged by the proper standards."). In my view, however, Gardner and Macklem fail fully to appreciate the fact that provocation and diminished capacity are \textit{partial} defenses. Instead, they tend to assimilate provocation to duress and diminished capacity to insanity, both of which (duress and insanity) are full excuses. Moreover, because they see duress as a respectable excuse and insanity as a shameful exemption, they tend naturally to see provocation as a respectable partial excuse and diminished capacity as a shameful partial exemption. Furthermore, Gardner and Macklem's criticism of diminished capacity as a shameful plea leaves them committed to the awkward proposition that battered women who kill their batterers and who claim diminished capacity are less than complete moral agents. \textit{See} Macklem \& Gardner, \textit{supra} note 22, at 827.

\textsuperscript{200} Nicola Lacey appears to take the same view. \textit{See} Lacey, \textit{supra} note 121, at 119 ("Exemptions should operate as total rather than partial defences . . . .").

\textsuperscript{201} \textit{See} Mackay \& Mitchell, \textit{Replacing Provocation}, \textit{supra} note 195, at 221 ("[T]here is no reason why the diminished defendant should be regarded as having less self-respect than his provoked counterpart.").

\textsuperscript{202} \textit{See supra} pp. 1692-1717.
provoked actor and the diminished actor kill, they do so because, while they
could and should have exercised control over their desire to kill, they failed
to do so. This failure renders each of them fairly liable to the law's
chastisement, although the punishment assigned to them is and should be
less than that assigned to the defiant killer. But chastisement should only be
a part, and perhaps only a small part, of the law's purpose when it responds
to provoked and diminished actors.

The punishment of provoked and diminished actors should also, if not
primarily, be seen as an occasion for reformation. It should be seen as an
occasion to offer the actor the opportunity to improve his prospects for
successfully exercising his capacity for self-control in the future. Here is
where the surface differences between the provoked actor and the
diminished actor may begin to make a difference.

Because the provoked actor suffers from no identifiable mental or
emotional abnormality, efforts to improve his prospects for the successful
exercise of his capacity for self-control might be described as a form of
character education. He should be obligated to learn various strategies and
techniques for the effective exercise of self-control, as well as to fortify his
will to deploy those strategies and techniques. Ideally, the provoked actor
will discharge this obligation gladly, committing himself to his own self-
reform and the hard work it requires. The goal is for the exercise of self-
control to become for him more or less a matter of natural disposition and
habit, and thus a part of his character. In other words, beyond chastising
him for his culpable failure of self-control, the provoked actor's punishment
should insist on his making an honest effort to leave the ranks of the akratic
(whose efforts to exercise self-control often fail) and to join instead the
ranks of the enkratic (whose efforts often succeed).

203. For a helpful discussion of weakness of will as a failure of character (as opposed to a
failure of rational action), see Thomas E. Hill, Jr., Weakness of Will and Character, 14 PHIL. TOPICS
No. 2, at 93 (1986).

204. The psychological literature supports the idea that an actor's will-power can over time
and with effort be strengthened. Consider this recent summary of the available research:

  Research has begun to suggest that the capacity for altering the self is a form of
  strength. Specifically, altering the self requires a mental or psychological exertion.
  Just as a muscle grows tired from exertion, the capacity for self-regulation becomes
  depleted after it is used. Also like a muscle, the capacity for self-regulation appears
to grow stronger through regular exercise.

CHRISTOPHER PETERSON & MARTIN E.P. SELIGMAN, CHARACTER STRENGTHS AND VIRTUES 501
(2004) (internal citations to psychological studies omitted).

205. The character education I have in mind would seem to fit comfortably within the
mission statement of a liberal state, being limited as it is with respect to both means and ends.
As to ends, the only virtue it aims to teach is the virtue of self-control, also known as continence
or temperance. See Duff, supra note 27, at 147 (discussing the distinction between virtue and
continence). As for means, it should proceed through education, not Clockwork Orange-like
indoctrination, and should thus exclude any technique, such as brainwashing or similar modes
of behavior modification, which violate an actor's autonomy.
In contrast, because the diminished actor does suffer from some identifiable mental or emotional abnormality, any effort to improve his prospects for the successful exercise of self-control should first take the form of some appropriate psychological treatment. The disorder at the root of his failed effort to exercise self-control should be addressed through appropriate and non-coercive means.

Of course, the point at which character education ends and psychological treatment begins is sometimes hard to identify. Just as the line between the provoked actor and the diminished actor can blur, so too can the line between the two different ways in which the law should respond to them. Are so-called anger management classes character education or psychological counseling? Yet whatever the label, the ultimate goal is the same. The punishment imposed on the diminished actor, just like that imposed on the provoked actor, should chastise him for his failure on the present occasion to successfully exercise his capacity for self-control, but it should also offer him the opportunity to improve the state of his psyche, thereby improving his prospects for the successful exercise of that capacity in the future.

The defiant actor displays contempt for the law and excessive pride in himself.206 In his mind, only the little people follow the law. Ideally, a defiant actor will see or come to see his punishment as part of a larger process through which he repents and atones for his offense, embracing his punishment as a symbolic means to repudiate the contempt expressed through his action. In other words, the defiant actor will ideally come to understand his punishment as a form of secular penance through which he expresses repentance and humility. Moreover, having endured this penance he will have earned forgiveness and the right to be readmitted to full membership in the community from which his crime has alienated him.

Provoked and diminished actors do not defy the law. Their actions display no contempt for the law, nor excessive pride in themselves. The vice displayed in the actions of the provoked and diminished actor is the lesser vice of intemperance or incontinence. Provoked and diminished actors fail to summon the self-control it takes to avoid violating the law. As such, these actors must embrace the burden imposed on them, not as a symbolic expression of the humility of a once-defiant will, but rather as an expression of their commitment to a course of self-reform, a course embarked upon in the hope that their efforts to exercise the virtue of self-control will meet with greater success in the future.

206. The ideas contained in this and the following paragraph are developed at greater length and in various contexts in Stephen P. Garvey, Is It Wrong To Commute Death Row? Retribution, Atonement, and Mercy, 82 N.C. L. REV. 1319 (2004); Garvey, supra note 163; Stephen P. Garvey, Restorative Justice, Punishment, and Atonement, 2003 UTAH L. REV. 303; Garvey, supra note 24.
Provocation is a venerable but besieged defense. It fights on one front against those who say it should be abolished. It fights on another against those who would portray it as a reward for virtue and a punishment for vice. In my view, the defense should be neither abolished nor so portrayed. Instead, it should be retained and recognized as one way in which the law separates those who defy its obligations from those who violate them out of culpable ignorance or weakness. In other words, it should be understood as a way to single out those “cases in which justice appear[s] to require a concession to human frailty.”

See sources cited supra note 32.

See Kahan & Nussbaum, supra note 23, at 366 (stating that the provocation doctrine “demands of people not only that they conform their conduct to a certain standard, but also that they shape their characters, and the quality of their emotions, in accordance with prevailing norms of reasonableness”).

Regina v. Smith, [2000] 4 All E.R. 289, 302 (H.L.) (opinion of Lord Hoffman) (emphasis added). As the Supreme Court has similarly noted,

The law, in recognition of the frailty of human nature, regards homicide committed under the influence of sudden heat of passion, or in hot blood produced by adequate cause, and before a reasonable time has elapsed for the blood to cool, as an offence of a less heinous character than murder.

Andersen v. United States, 170 U.S. 481, 510 (1898).