HUMAN FRAILTY: THE STRUCTURE AND FOUNDATION OF THE PROVOCATION DEFENSE

Gabriel S. Mendlow
Yale Law School

INTRODUCTION

Legal theorists commonly distinguish between theories that interpret the law and theories that attempt to justify it. What the law says is one question; why (or whether) the law is justified is another. But there are areas of law where these two questions have the same answer, because there are areas of law where a sound interpretation of the doctrine reveals the doctrine to be sound. One such area is the provocation defense, which reduces an intentional homicide from murder to voluntary manslaughter where the defendant killed his provoker in the heat of passion. A sound interpretation of the provocation defense reveals a normative structure that wears its justification on its sleeve. In this area, to interpret the law correctly is to justify it—or so I hope my interpretation will reveal.

The oldest and least controversial prototype for the provocation defense is the defendant who kills in the middle of a fight or in retaliation for an attack. Nearly everyone agrees that this defendant deserves some measure of leniency. A much less compelling prototype is the defendant who kills his spouse or his spouse’s lover after discovering the two of them in bed together. This prototype seems to indulge domestic violence and to credit the archaic notion that a man’s wife is in some sense his property. Theorists have accordingly questioned whether the doctrine of provocation should remain part of our criminal law.\(^1\) I will argue that it should. Properly interpreted, the doctrine of provocation is justified; properly applied, it does not sanction misogyny.

Theorists have not interpreted the doctrine of provocation properly because they have not really interpreted the doctrine. Instead, they have interpreted the classical gloss on the doctrine:

\[
D \text{ is guilty of voluntary manslaughter rather than murder if she intentionally killed } V \text{ in the following circumstances: (1) when she killed } V \text{ she was in a state of passion [i.e., heat of passion]; (2) the}
\]

---

\(^1\) E.g., Jeremy Horder, Provocation and Responsibility (1992).
passion was caused by adequate provocation; and (3) her homicidal acts occurred before she had a reasonable time to cool off.\textsuperscript{2}

Echoed by codes, courts, and commentators, the classical gloss on the provocation defense conceals the doctrine’s true complexity—complexity that I aim to uncover and expound, first by reconstructing the doctrine in a way that lays bare its moral structure, and then by examining that structure’s psychological foundation.

A chief sin of the classical glass is its coarse-grained psychology. Under the label “heat of passion,” the classical gloss conflates two psychological elements that the law of provocation fundamentally distinguishes: the defendant’s emotional reaction to being provoked and her consequent volitional impairment or loss of self-control. Only if we recognize these psychological elements as distinct can we understand why provocation mitigates. Properly interpreted, the provocation defense deems certain killers less than fully blameworthy where their homicidal conduct resulted from an understandable loss of self-control that resulted in turn from a flash of understandable anger. Together, anger and loss of self-control lessen a killer’s guilt. Neither does so alone. Because the leading theories of the provocation defense mistake the classical gloss for a true statement of the doctrine, they distinguish inadequately between anger and volitional impairment and therefore fall short of illuminating why provocation mitigates.

Part I of the Article rejects the classical gloss, arguing that the doctrine of provocation is best understood in terms a six-part formula. The six-part formula submits the defendant’s anger and volitional impairment to successive rounds of normative scrutiny. Both rounds invoke a moral category largely absent from legal theory but covertly present in the law: the category of the morally understandable. To deem someone’s anger or volitional impairment morally understandable is not to deem it justifiable or excusable, but rather to deem it partly excusable, partly blameworthy. Captivated by the idea that blameworthiness is all-or-nothing—that blame and excuse are a binary pair—legal theorists have failed to account for what all of us have long known: even when our shortcomings are not altogether excusable, they may not be altogether blameworthy, either.

Notwithstanding appearances to the contrary, no aspect of the six-part formula is a departure from the common law. As I show, the traditional doctrine of provocation applies the so-called “reasonable person standard” not once but twice: first it appraises the defendant’s emotional reaction, then it appraises her loss of self-control. Moreover, the traditional doctrine assesses these aspects of the defendant’s psychological condition in terms of whether they are morally understandable—not in terms of whether they are justifiable or unjustifiable, excusable or blameworthy. If we appreciate these subtleties in the traditional doctrine, we will see that the established theoretical categories of justification, excuse, and blame are not up to the task of explaining why the provoked killer should be eligible for mitigation. In short, doctrinal clarity may lead to moral insight.

Doctrinal clarity may also lead to psychological insight. Where Part I of the Article elaborates provocation’s moral structure, Part II puts that structure on a deeper theoretical foundation, a philosophical conception of human motivation. The key challenge is to articulate a theory of the provoked defendant’s retaliatory anger that makes it intelligible how such anger could both (i) be morally understandable in light of the victim’s prior conduct and (ii) be the cause of morally understandable loss of self-control. I meet this challenge by offering a novel theory of affective desire. This theory is preferable to the leading philosophical conceptions of desire and emotion, which (as I argue) cannot accommodate the phenomenon of morally understandable loss of self-control. The upshot is that an adequate theory of the provocation defense demands a new theory of human motivation.

I. THE DOCTRINE AND ITS MORAL STRUCTURE

A. The Classical Gloss: Leading Theories

Contemporary criminal law scholarship offers two main theories of the classical gloss on the provocation defense. These two theories yield contrasting answers to the question why a killer’s passion sometimes lessens the gravity of his crime.3 The volitional theory focuses on the

---

3 A third approach justifies the provocation defense on utilitarian grounds, holding that hot-blooded killers are either less dangerous or less able to be deterred than cold-blooded ones. See, e.g., Jeremy Bentham, An Introduction to the Principles of Morals and Legislation; Jerome Michael & Herbert Wechsler, A Rationale of the Law of Homicide II, 37 Colum. L. Rev. 1261 (1937); Gary S.
passion’s intensity, assessing the passion’s relationship to the killer’s subsequent loss of self-control. A key question for the volitional theory is whether the killer’s loss of self-control was reasonable given the intensity of his passion.\(^4\) The \textit{evaluative theory} focuses not on the passion’s intensity but on its propriety, assessing the passion’s relationship to the victim’s antecedent act of provocation. A key question for the evaluative theory is whether the killer’s passion was reasonable given the nature of the victim’s provocation.\(^5\)

I will argue in a moment that each of these theories contains a seed of truth; specifically, that the provocation defense is best understood as assessing two things together: the passion’s propriety as well as its intensity. But first I want to consider the leading theories separately and see how each explains the classical gloss on its own.

The volitional theory roots mitigation in volitional impairment: a hot-blooded killer is less blameworthy than a cold-blooded one because

---

\(^4\) George P. Fletcher, Rethinking the Criminal Law 242-43 (1978) (“Though provocation does not enjoy a standard definition, the issue is generally understood to be whether the defendant was substantially impaired in his self-control at the time of the killing…. The issue is plainly normative in the sense that the homicide is not mitigated to manslaughter by a mere factual showing that the slayer was provoked. He must be provoked under circumstances and to such a degree that he is not expected completely to control himself.”); Joshua Dressler, Rethinking Heat of Passion: A Defense in Search of a Rationale, 73 J. Crim. L. & Criminology 421 (1982); Joshua Dressler, When “Heterosexual” Men Kill “Homosexual” Men: Reflections on Provocation Law, Sexual Advances, and the “Reasonable Man” Standard, 85 J. Crim. L. & Criminology 726 (1995); Joshua Dressler, Why Keep the Provocation Defense?: Some Reflections on a Difficult Subject, 86 Minn. L. Rev. 959 (2002).

\(^5\) Dan M. Kahan and Martha C. Nussbaum, Two Conceptions of Emotion in Criminal Law, 96 Colum. L. Rev. 269 (1996); see also Samuel H. Pillsbury, Emotional Justice: Moralizing the Passions of Criminal Punishment, 74 Cornell L. Rev. 655, 678 (1989) (“The doctrine [of provocation] is available only for those situations where the provocation is reasonable—where we can understand and morally sympathize with, if not excuse, the motivation.”).
the hot-blooded killer has far more trouble exercising self-control. In the words of Joshua Dressler, the leading volitional theorist,

"[the true reason for the law’s “concession to human weakness”—the reason why, if A kills P in sudden rage at his actions, the law will likely allow A to argue that the jury should reduce the homicide to manslaughter—is that the homicide is the result of an understandable and excusable loss of self-control arising from his anger. Common experience teaches that, at some point, anger becomes so intense that people find it extremely difficult to control themselves and respond constructively, rather than violently, to the anger-producing stimulus. Therefore, when A kills P because his reason is “disturbed or obscured by passion to an extent which might render ordinary men, of fair average disposition, liable to act rashly or without due deliberation or reflection, and from passion, rather than judgment,” he is less to blame than if he killed P while he was calm. This is because it is harder for A to control his actions when he is angry than when he is calm." 

Rooting mitigation in volitional impairment, the volitional theory interprets each element of the classical gloss as an indicator of loss of self-control. A provocation is *adequate* if it is sufficient to cause a person endowed with ordinary powers of self-restraint to have serious difficulty controlling herself; a person acts in the *heat of passion* if she kills under the influence of an emotion so strong that it “dominates [her] volition”; and a person kills in the absence of *reasonable cooling time* if she acts before a person of ordinary self-restraint would have regained control, so that “the killing [can be] seen as an outgrowth of the provocative event,

---

8 Dressler, When “Heterosexual” Men Kill “Homosexual” Men, supra note _, at 747-48 (some internal citations omitted).
9 2 Wayne R. LaFave, Substantive Criminal Law 491 (2003) (defining an adequate provocation as “a provocation which would cause a reasonable man to lose his normal self-control”); cf. Dressler, Rethinking Heat of Passion, supra note _, at 466-67 (“[If the provocation] is so great that the ordinarily law-abiding person would be expected to lose self-control to the extent that he could not help but act violently, yet he would still have sufficient self-control so that he could avoid using force likely to cause death or great bodily harm in response to the provocation, then we are saying that the actor’s moral blameworthiness is found not in his violent response, but in his homicidal violent response”).
10 Smith v. State, 3 So. 551, 552 (Ala. 1888).
[instead of] as an independent act for which [she] is fully accountable.'\textsuperscript{11} According to the volitional theory, the reason why provocation \textit{mitigates} instead of fully \textit{exculpating} is that the provoked killer does not wholly lack the capacity to control herself: however difficult it might be for an "adequately" provoked person to refrain from acting on her violent passion, it is not altogether impossible.\textsuperscript{12}

The evaluative theory shifts the focus away from volitional impairment and interprets each element of the classical gloss as an indicator not of a passion that is especially intense but of a passion that is especially appropriate. According to the evaluative theory, a provocation is \textit{adequate} not (necessarily) when it induces a passion that overcomes one's self-control but instead when it is an affront in response to which one rightly becomes impassioned. A defendant acts in the \textit{heat of passion} not (necessarily) when her emotion is so strong that it "dominates [her] volition";\textsuperscript{13} rather, a person acts in the heat of passion when (in the words of the leading evaluative theorists, Dan Kahan and Martha Nussbaum) the defendant's emotion embodies "an appropriate valuation of the good . . . that is threatened by the victim's wrongful provocation."\textsuperscript{14} And a person kills in the absence of \textit{reasonable cooling time} when she acts while it is (still) morally appropriate to be impassioned. According to the evaluative theory, the reason why provocation mitigates instead of fully exculpating is that the defendant's passion embodies a reasonable but imperfect valuation of the good that is threatened by the victim's wrongful provocation: the defendant is right to value the threatened good

\textsuperscript{11} Fletcher, Rethinking Criminal Law, supra note _, at 244.
\textsuperscript{12} Dressler, Why Keep the Provocation Defense?, supra note _, 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 \textit{wholly} incapable of controlling or channeling his anger.").
\textsuperscript{13} Smith, 3 So. at 552.
\textsuperscript{14} Kahan and Nussbaum, Two Conceptions of Emotion, supra note _, at 315 (1996) ("[W]ithout rage or anger, the defendant's acts would express nothing more than an inappropriately low valuation of the victim's life. The common-law authorities tend to justify the 'heat of passion' requirement in exactly this way, explaining that a killing without appropriate emotion necessarily reflects 'wickedness of heart or cruelty.'" (quoting Maher v. People, 10 Mich. 212, 220 (1862))) (citing State v. Gounagias, 153 P. 9, 12 (Wash. 1915)).
but wrong to value it so much more than she evidently values the victim’s life.\textsuperscript{15}

Against the volitional theory, evaluative theorists emphasize aspects of the common-law doctrine that reveal volitional impairment to be insufficient to downgrade a homicide from murder to voluntary manslaughter. Evaluative theorists point out what common-law courts have long held, that adequacy of provocation goes not to a passion’s intensity but to its moral quality. A “provocation [cannot] be held sufficient or reasonable [simply] because . . . a state of excitement [i.e., an intense passion] has followed from it; for then, by habitual and long continued indulgence of evil passions, a bad man might acquire a claim to mitigation which would not be available to better men, and on account of that very wickedness of heart which, in itself, constitutes an aggravation both in morals and in law.”\textsuperscript{16} Were the doctrine of provocation indifferent to a passion’s moral quality, the doctrine “would make evil the palliative of crime, and vice would become an excuse for its own fruits.”\textsuperscript{17} No matter how intense a defendant’s passion is, it will not mitigate his guilt if it springs from “a heart . . . devoid of social duty, that is, reckless of the rights and lives of others, and fatally bent on

\begin{quote}
\textsuperscript{15} Kahan and Nussbaum, Two Conceptions of Emotion, see supra note __, at 313 (“To make this concrete, imagine a woman who kills a man in anger after discovering that he has sexually abused the woman’s young daughter. From the evaluative perspective, one would say that her emotion embodies appraisals of mixed quality. She has appraised her circumstances in a way correctly, since her anger reflects her appropriate valuation of her daughter’s well-being; but in a way she has also appraised them wrongly, since she should not have thought that this good was all-important, taking precedence over all other considerations, including the value of the man’s life and the importance of lawful resolution of disputes. Her judgment may have been distorted because she harbored a skewed relative valuation of the different goods involved or, more subtly, because she focused so intensely on one of them that other relevant considerations were temporarily eclipsed from view. But either way, our assessment of her behavior is likely to be complex; her emotional motivation is reasonable, but imperfect. The mitigating consequence of voluntary manslaughter captures the complexity of this assessment.” (internal citations omitted)).


\textsuperscript{17} Rivers v. State, 78 So. 343, 345 (Fla. 1918) (cited by Kahan and Nussbaum, Two Conceptions of Emotion, supra note __, at 307).
\end{quote}
taking life. . . .”\textsuperscript{18} For a killing to qualify as manslaughter rather than murder, the “killing [must] proceed [not] from a bad or corrupt heart, [but] rather from the infirmity of passion to which even good men are subject.”\textsuperscript{19} So goes the doctrinal argument against the volitional theory.

Now evaluative theorists are undoubtedly correct to balk at the volitional theory’s single-minded focus on loss of self-control, since the common-law authorities that I just quoted make plain that mitigation demands more than mere volitional impairment. Volitional impairment is indeed insufficient. But evaluative theorists are wrong to suppose that volitional impairment is also unnecessary. As we will see, the provocation defense cares deeply about loss of self-control. This is something Kahan and Nussbaum deny, claiming that the common-law authorities “do not insist that the defendant’s passion be of any particular intensity—much less of volition-distorting intensity. To justify a verdict of voluntary manslaughter, the jury need not find that the killing followed from the defendant’s rage with ‘the certainty that physical effects follow from physical causes.’”\textsuperscript{20} Here Kahan and Nussbaum quote the Michigan Supreme Court’s opinion in \textit{Maher v. People}, a landmark provocation case. Just a few paragraphs earlier in the same opinion, however, the court explicitly specified “the extent to which the passions are required to be aroused.”\textsuperscript{21} Far from saying that the defendant’s passion need not be of any particular intensity, the court held that the passion must be intense enough to “[disturb] the control of reason.”\textsuperscript{22} Granted, the court also observed that the defendant’s passion must not be too intense: “It will not do to hold that reason should be entirely dethroned, or overpowered by passion so as to destroy intelligent volition. . . . Such a degree of mental disturbance would be equivalent to utter insanity, and, if the result of adequate provocation, would render the perpetrator morally innocent.”\textsuperscript{23} But these remarks entail only that the passion’s intensity has an \textit{upper} bound, not that it lacks a \textit{lower} bound.

\begin{thebibliography}{10}
\bibitem{18} State v. Cook, 3 Ohio Dec. Reprint 142, 147 (1859).
\bibitem{19} Id. at 146 (cited by Kahan and Nussbaum, Two Conceptions of Emotion, supra note \_\_ at 307).
\bibitem{20} Kahan and Nussbaum, Two Conceptions of Emotion, supra note \_\_, at 316 (internal citation omitted) (quoting \textit{Maher}, 10 Mich. at 220).
\bibitem{21} \textit{Maher}, 10 Mich. at 220.
\bibitem{22} Id. at 219.
\bibitem{23} Id. at 220 (internal citations omitted).
\end{thebibliography}
B. The Six-Part Formula: Overview

If the case law is any guide, the provocation defense takes heed of a passion’s intensity as well as its moral quality. But why both? Given that provocation mitigates instead of fully exculpating, shouldn’t it be enough that the provoked defendant act from an emotion sufficiently intense to impair her self-control, whether or not that emotion is morally appropriate? Alternatively, shouldn’t it be enough that the provoked defendant act from an emotion that is morally appropriate, whether or not that emotion is sufficiently intense to impair her self-control? And shouldn’t the defendant be fully excused if she acts from an emotion that is at once morally appropriate and sufficiently intense to impair her self-control?

We will be in a better position to answer these questions about provocation’s moral basis if we gain a clearer view of its doctrinal structure. To that end, I propose that we jettison the classical gloss in favor of a six-part formula. As our exploration of the common-law treatment of provocation has already revealed, there is a good deal more going on in the law than the classical three-part gloss can comfortably allow.

The six-part structure I am about to delineate will look somewhat different from (though not completely unlike) the classical formula. In that respect, my project is reformist: if the formula I propose is sound, legislatures might revise their criminal codes, courts might change the way they talk about provocation and voluntary manslaughter (both in their instructions to juries and in their written opinions), and we might consider new ways of dividing the deliberative burden between judges and juries. But the formula I propose is not fundamentally novel. It has an analytical structure that is already present in the law, albeit covertly. Once this analytical structure is fully in view, we will able to see why the law of provocation makes moral sense. We may also be able to see how the law might be applied more consistently.

The first step toward achieving doctrinal clarity is to distinguish the provocation defense’s factual components from its normative ones. Part of what makes the three-part gloss misleadingly spare is that at least two of the parts combine the factual and the normative. When we determine whether the defendant killed upon “adequate provocation,” we are answering both the factual question whether there was a provocation and the normative question whether the provocation was adequate. Inquiring into a provocation’s adequacy is a matter of asking whether the
provocation would lead a reasonable person to become impassioned (the
evaluative interpretation) or whether the provocation would lead a
reasonable person to lose self-control (the volitional interpretation). Both
questions are normative because the concept of a reasonable person is
fundamentally a normative concept: it is the concept of how a person
ought to be constituted, in particular, how a person ought to be
constituted psychologically—with respect to her emotional propensities
or her powers of self-regulation or both. Furthermore, when we
determine whether the defendant acted in the absence of reasonable
“cooling time,” we are answering not only the factual question whether
the defendant’s passion had (or had not) subsided before the killing but
also the normative question whether the amount of time that elapsed
before the killing was sufficient—sufficient either for a reasonable
person to regain self-control (the volitional interpretation) or for a moral
person to stop seething about the affront and “let go” (the evaluative
interpretation). 24 Again, both questions are inescapably normative, given
the obviously normative status of the concept of a reasonable or moral
person. Arguably, there are also separate factual and normative
components to “heat of passion”—at least according to the evaluative
theory, which holds that the defendant’s passion must in fact be aroused
and that the passion must also embody an appropriate valuation of the
good that is threatened by the victim’s wrongful provocation.

With a view to keeping the factual separate from the normative, I
propose the following formulation of the provocation defense:

**In a prosecution for murder, it is a defense that the defendant
killed as a result of**

(1) a *loss of self-control* caused by
(2) *anger* aroused by

---

24 Kahan and Nussbaum, Two Conceptions of Emotion, supra note __, at 318
(“[I]f the cuckold continues to be obsessively angry for days, weeks, months, or
even years, then we will regard his view of what’s important in life as skewed.
Unshakable rage reveals that he values something—honor, control—too much.
It also reveals that he values other things that matter—like the life of the victim
and the lawful resolution of disputes—too little. If a person has good character,
he won’t stop valuing fidelity as time passes, but he will take the steps
necessary—perhaps divorcing his wife, or possibly notifying authorities of the
transgression of the offending man—to restore “tranquility of . . . judgment” in a
way that reflects an appropriate valuation of all the goods and interests at stake.”
(quoting People v. Ashland, 128 P. 798, 802 (Cal. Ct. App. 1912))).
(3) the victim’s conduct,
provided that, as determined by the finder of fact,
(4) the defendant’s loss of self-control was morally understandable
given the nature and intensity of her anger and
(5) the nature and intensity of the defendant’s anger were morally
understandable given the victim’s conduct, and,
provided that, as determined by the finder of law,
(6) the victim’s conduct constituted a legally cognizable
provocation.

Putting everything together in chronological order, D’s intentional killing
of V was manslaughter rather than murder if V’s provocative conduct
[(3) and (6)] aroused in D a state of understandable anger [(2) and (5)],
which caused D understandably to lose self-control [(1) and (4)] and kill
V in retaliation. Elements (1)-(3) are factual: they are events that must
occur if a person’s homicidal behavior is to fall within the purview of the
provocation defense. Elements (4)-(6) are normative: they are tests that
the events described by (1)-(3) must pass if a plea of provocation is to
succeed. On the factual side, the critical departures from the classical
gloss are elements (1) and (2): unlike the classical gloss, which contains
only one explicitly psychological element—heat of passion—the six-part
formula treats emotion and volitional impairment as distinct. On the
normative side, the critical departures are elements (4) and (5): the six-
part formula bifurcates the ‘reasonable person standard,’ subjecting
the defendant’s emotion and subsequent volitional impairment to separate
scrutiny. As I will now argue, all of these departures are departures
merely from the classical gloss, not from the common law itself. Each
element of the six-part formula is part of the traditional doctrine of
provocation, whether or not explicitly.

The point to all this is more than mere doctrinal precision—although
there is a point to that, too. The deeper point is that the doctrine of
provocation, properly expounded, wears its justification on its sleeve. To
know what the doctrine says is to know why it is justified. So let us have
a closer look at what the doctrine says.

C. The Factual Story

I. The victim’s conduct [(3)]. — Critical to the common-law
provocation defense is the idea that the victim must have done
something. This is a major point of contrast with the Model Penal Code,
which grants a similar defense to anyone who commits a homicide
“under the influence of extreme mental or emotional disturbance for
which there is reasonable explanation or excuse." Under the Model Penal Code, the defendant need not have been provoked by the victim; indeed, the defendant need not have been provoked at all.

2. Anger aroused by the victim’s conduct [(2)]. — The provocation defense requires that the defendant’s anger have a particular origin and a particular trajectory. The defendant’s anger must have originated as a response to the victim’s conduct; where the defendant was already angry at the victim, the provocation defense is inapplicable. Furthermore, the provocation defense requires not just that the defendant kill in anger but that he kill from anger. Anger must be what impels the defendant to kill. It must be his primary motivation.

No emotion other than anger will render a defendant eligible for the traditional provocation defense. Here, again, the traditional defense is considerably narrower than the Model Penal Code’s defense of extreme mental or emotional disturbance, which permits consideration of such emotions as fear and despair. Though courts sometimes claim that the

---

25 American Law Inst., supra note __, §210.3(1)(b) (emphasis added).
26 E.g., People v. Casassa, 404 N.E.2d 1310 (N.Y. 1980) (permitting the defense where there was no provocation).
27 See Andrew Ashworth, The Doctrine of Provocation, 35 Cambridge L.J. 292, 294 (1976) (“In early cases . . . [w]here . . . there was a finding of ‘precedent malice,’ the provocation was deprived of all legal effect. In the leading case of Maddy [late seventeenth century], where a husband returned home to find his wife in the act of adultery and struck the man on the head with a joint-stool, the jury were asked to find whether Maddy had precedent malice, in the form of a prior determination to take revenge. So even where the provocation was of the highest degree, the element of sudden passion had to be established. If he had known of the association and had declared his intention to take revenge, not even discovery in flagrante delicto would have reduced his offence from murder to manslaughter.” (internal citation omitted)).
28 People v. Steele, 47 P.3d 225, 239-40 (Cal. 2002) (“The defendant must actually, subjectively, kill under the heat of passion.”).
29 E.g., State v. Beegle, 425 S.E.2d 823, 828 (W.Va. 1992) (voluntary manslaughter is by definition homicide committed in heat of passion; the defendant is not entitled to a voluntary manslaughter instruction where he claims to have shot the victim out of fear rather than in anger).
30 State v. Elliot, 411 A.2d 3 (Conn. 1979) (defendant was very afraid of his brother); State v. Zdanis, 438 A.2d 696 (Conn. 1980) (defendant was very upset about the fact that his young niece had cancer).
traditional doctrine permits consideration of fear—some even claim that the defense permits consideration of any violent and intense emotion—these remarks are seemingly always dicta. I am aware of no case falling under the common-law provocation defense in which the operative emotion was in fact anything other than anger. To be sure, the common law sometimes reduces a homicide from murder to voluntary manslaughter on account of a defendant’s fear, but it does so under the doctrine of imperfect self-defense, not provocation.

The provocation defense’s restriction to anger is neither accidental nor arbitrary. The inner logic of anger is hand in glove with the moral structure of the provocation defense, which is centrally about retaliation against an apparent wrongdoer mediated by an emotion. Anger is perfectly suited to being this emotion since it often involves (a) the perception that one has been wronged and (b) the desire to retaliate

31 See, e.g., Barrett v. Commonwealth, 341 S.E.2d 190, 192 (Va. 1986) (“Heat of passion . . . may be founded upon rage, fear, or a combination of both . . .”); McBroom v. State, 530 N.E.2d 725, 728 (Ind. 1988) (“Sudden heat is anger, rage, resentment, or terror sufficient to obscure the reason of an ordinary man . . .”); Commonwealth v. Harris, 372 A.2d 757, 759 (Pa. 1977) (“Passion includes any emotions of the mind known as anger, rage, sudden resentment or terror, rendering the mind incapable of cool reflection . . .”).

32 People v. Breverman, 960 P.2d 1094, 1107-08 (Cal. 1998) (“[T]he passion aroused need not be anger or rage, but can be any ‘violent, intense, high-wrought or enthusiastic emotion’” (internal citation omitted)).

33 On imperfect self-defense, see generally Paul Robinson, Criminal Law Defenses §132.50 (2010) (“An instruction on imperfect self-defense should be given where a defendant reasonably believes it necessary to kill the deceased to save himself from death or great bodily harm even if defendant (1) might have brought on the difficulty, provided he did so without murderous intent, and (2) might have used excessive force. If the belief in the need to defend subjectively exists but is objectively unreasonable, there is imperfect self-defense, i.e., the defendant is deemed to have acted without malice and cannot be convicted of murder, but can be convicted of manslaughter.”).

34 Cf. Reid Griffith Fontaine, Adequate (Non)Provocation and Heat of Passion as Excuse Not Justification, 43 U. Mich. J.L. Reform 27, 48 (2009) (“Heat of passion reflects a mediate path by which there is an interpretation of provocation (cognition) that causes overwhelming anger (emotion), which, in turn, leads to a reactive killing (behavior).”).
against the apparent wrongdoer. The apparent wrongdoer is anger’s object twice over, that is, perceptually as well as desideratively—just as the provocation defense requires that the apparent wrongdoer be the provoker as well as the victim. No other emotion has an inner logic that aligns so neatly with provocation’s moral structure. (I will say a great deal more about the inner logic of retaliatory anger in Part II.)

3. Loss of self-control caused by defendant’s anger (1). — The provocation defense requires not just that the defendant’s anger motivate his homicidal conduct but that it motivate his conduct in a particular way: by means of a loss of self-control. It is not enough that the defendant kill in and from anger; the defendant’s anger must impair his volition. As the Minnesota Supreme Court explained in State v. Quick, “if a defendant is in the heat of passion, his reason [must] be clouded and his willpower weakened. Anger alone is not sufficient. . . .” These words are not mere dicta. In Quick, the court upheld the defendant’s

---

35 Timothy Macklem and John Gardner, Provocation and Pluralism, 64 Mod. L. Rev. 815, 820 (2001) (“It is of the essence of anger that there is a (supposed) wrongdoer against whom the anger is directed.”).

36 Stephen P. Garvey, Passion’s Puzzle, 90 Iowa L. Rev. 1677, 1687 n.34 (“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.”).


38 Attorney General’s References Nos. 74, 95, and 118 of 2002 (Darren Suratan, Leslie Humes, and Mark Wilkinson) [2003] 2. Cr. App. R. (S.) 42, para. 24 (“At the time of the killing, [the offender must have] lost his self-control. Mere loss of temper or jealous rage [is] not sufficient.”).

39 State v. Quick, 659 N.W.2d 701, 711 (Minn. 2003) (internal citation omitted).
conviction of first-degree murder even as it credited evidence that the defendant was extremely angry.\textsuperscript{40}

Planned revenge-killings present the clearest cases in which defendants are ineligible for the provocation defense because they kill from anger without being volitionally impaired.\textsuperscript{41} In \textit{Regina v. Ibrams}, \textsuperscript{42} the English Court of Appeal found insufficient evidence of loss of self-control where the defendant had planned and perpetrated a revenge killing. The court reached this conclusion despite sympathizing with the defendant’s plight and crediting evidence that the victim’s repeated bullying had understandably aroused the defendant to a state of great emotional intensity.\textsuperscript{43} The lesson is that the law of provocation

\begin{footnotesize}
\begin{enumerate}
\item Id. at 706-12.
\item Planned revenge-killings have long been held ineligible for the provocation defense. See Ashworth, \textit{Principles of Criminal Law}, supra note _, at 252 (“[F]or centuries courts have drawn a distinction between revenge killings and others where [the defendant] was not ‘the master of his own understanding’ and where there had not been ‘time for the blood to cool and for reason to resume its seat.’” (quoting Hayward, [1833] 6 C.P. 157)); see also Felix v. State, 18 Ala. 720 (1851) (“The law carefully distinguishes between a sudden transport of passion, which springs instantaneously from what it allows as a sufficient provocation and which prompts to an immediate act of violence, and a purpose of revenge, which usually follows such passion. In the first case, in condescension to the frailty of our nature, the law allows the provocation to extenuate a homicide committed at the instant from murder to manslaughter. In the other, the provocation, furnishing an incentive to revenge, so far from extenuating the crime, is a circumstance to be looked to as evidence of malice, and especially would this be so, if the prisoner, in consequence of the provocation, had made threats against the life of the deceased.”); State v. Yanz, 50 A. 37, 39 (Conn. 1901) (“It should . . . be remembered that to call for the milder punishment the killing must be in fact the result of a sudden rage, difficult for the ordinary man to control, directly induced by a grievous injury. If in fact it is the result of the cruel spirit of revenge that must have life for a wrong, it is murder, no matter what the provocation may be.”); State v. Yarborough, 18 P. 474, 479 (Kan. 1888) (holding that defendant was ineligible for the provocation defense because he had “had ample time to exercise control or mastery over his passions, but nursed his wrath, and committed the fatal act with express malice”).
\item Id. at 159 ([G]ross bullying and terrorising . . . had almost certainly impaired the [defendant’s] judgment . . . but that impairment . . . is not the same as loss of self-control. Impairment of judgment led [the defendant and his)
\end{enumerate}
\end{footnotesize}
distinguishes between emotion and volitional impairment: both are necessary; hence, neither is sufficient.

In sum, the factual story has three main parts—the victim’s conduct [(3)], the defendant’s emotion [(2)], and the defendant’s loss of self-control [(1)]. These parts are causally linked: the victim must engage in conduct that arouses the defendant to a state of intense anger, and this anger must then cause the defendant to lose control and kill the victim.

Where the six-part formula distinguishes cleanly between the doctrine’s various factual components, the classical gloss tends to conflate them. Under the heading of “heat of passion,” the classical gloss emphasizes the element of emotion [(2)] at the expense of the equally important element of volitional impairment [(1)]—much the inverse of how Section 3 of the English Homicide Act of 1957 (defining the provocation defense) draws admirable attention to volitional impairment while neglecting the precipitating emotion:

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 . . .

The English statute is wrong not to mention the defendant’s emotion, just as the classical gloss is wrong not to mention volitional impairment. As the case law reveals, each of these two psychic elements—emotion and volitional impairment—is a separate component of the provocation defense.

We will soon see that the significance of their separateness is not merely descriptive; it is also normative. Critical to the defense’s moral structure is a bifurcated reasonable person standard, a standard that assesses the defendant’s emotion separately from his state of volitional impairment.
D. The Normative Tests

The six-part formula submits each factual element of the provocation defense to a normative test: the victim’s conduct [(3)] must constitute a legally cognizable provocation [(6)] (the *provocation requirement*); the defendant’s emotion [(2)] must be morally understandable in light of the victim’s conduct [(5)] (the *emotional requirement*); and the defendant’s subsequent loss of self-control [(1)] must be morally understandable in light of the nature and intensity of his emotion [(4)] (the *volitional requirement*). The provocation requirement imposes a limiting condition: it rules out certain circumstances in which a defendant’s psychological condition would otherwise diminish her culpability. The emotional and volitional requirements are the doctrine’s moral core. They explain why a killer’s psychological condition can diminish her culpability. They explain the sense in which the provocation defense involves “a concession to the frailty of man, a recognition that the average person can understandably react violently to a sufficient wrong and hence [that] some lesser punishment is appropriate.”

I. The Emotional and Volitional Requirements [(5) and (4)]. — A natural place to search for insight into the provocation defense’s moral core is the jurisprudence of the so-called “reasonable person” standard. Unfortunately, courts have done little to illuminate the “reasonable person.” This ideal person is variously described as “ordinary, reasonable, just and reasonable, ordinary and reasonable, ordinarily reasonable, ordinarily prudent, average, of fair average mind, or [as] an ordinary person of average disposition.” Not only are these terms non-equivalent, courts have mostly left them undefined. So our guiding light cannot simply be the doctrine’s jurisprudence. It must also be the doctrine’s effect: mitigation. The question, then, is this: what normative standard must a provoked killer’s psychological condition meet in order for him to be culpable, but not as culpable as a murderer? This question is potentially two questions, since the doctrine of provocation distinguishes between two aspects of the provoked killer’s psychological condition—his anger and his subsequent loss of self-control. We must

---


45 Joshua Dressler, Rethinking Heat of Passion, supra note _, at 432-33 (internal citations omitted).
accordingly define two normative inquiries: one that assesses the killer’s anger and another that assesses his loss of self-control.

I have indicated that these two normative inquiries ultimately apply the same standard, the standard of what is morally understandable. As a normative category, the understandability arises less often in law than does the reasonable. Moreover, the understandable is largely absent from criminal law theory, where the established moral categories are justification, excuse, and blame. (An action is justified if it is not wrong or criminal; an action is excused if, though unjustified, it is not blameworthy or culpable; and an action is blameworthy or culpable if it is neither justified nor excused.) No theorist suggests that we understand the provocation defense as a full-on justification or excuse, and for good reason: the defense does not fully exculpate. Rather, theorists typically apply the concepts of full-on justification and excuse not to the defense itself but to the defense’s mental elements. For example, theorists

---

46 Criminal law theorists do sometimes use the term “understandable,” but they typically do so in order to characterize one of the established moral categories—specifically, the excusable—not in order to name a separate category. Claire Finkelstein, for example, describes a duress defendant’s behavior as “understandable,” where her purpose is to emphasize that the reason why the defendant’s behavior is (fully) excusable is not that it was involuntary but that it flowed from a good character. Duress: A Philosophical Account of the Defense in Law, 37 Ariz. L. Rev. 251, 251 (1995) (“[In pleading duress] we accept responsibility for the deed, admit that it was bad, but argue that our behavior was understandable under the circumstances and that we should therefore not be punished.”). Instead of using the term “understandable” (as Finkelstein does) to describe the excusable, I am using term “understandable” to label a separate category. Moreover, I am suggesting that we conceive the law as applying this category primarily to the defendant’s psychological condition, and only derivatively to her behavior.

47 Robinson, Criminal Law Defenses, supra note _, §24(a) (“The harm caused by . . . justified behavior remains a legally recognized harm that is to be avoided whenever possible. Under the special justifying circumstances, however, that harm is outweighed by the need to avoid an even greater harm or to further a greater societal interest.”).

48 Excuses “exculpate even though the elements of the offense are satisfied. Excuses admit that the deed may be wrong, but excuse the actor because conditions suggest [either] that he is not responsible for his deed” or that, even if he is in some sense responsible, he is nevertheless not blameworthy. Id. §25(a).
propose that the law mitigates a provoked killer’s punishment because
his anger was justified, or because his loss of self-control was excusable.

(a) Justification, Excuse, and Blame

If we survey the full range of possibilities afforded by the categories
of justification, excuse, and blame, we will arrive at nine different ways
of understanding what it takes for a provoked killer’s psychic state to
satisfy the emotional and volitional requirements:

(i) Justifiable anger induces justifiable loss of self-control
(ii) Justifiable anger induces excusable loss of self-control
(iii) Justifiable anger induces blameworthy loss of self-control
(iv) Excusable anger induces justifiable loss of self-control
(v) Excusable anger induces excusable loss of self-control
(vi) Excusable anger induces blameworthy loss of self-control
(vii) Blameworthy anger induces justifiable loss of self-control
(viii) Blameworthy anger induces excusable loss of self-control
(ix) Blameworthy anger induces blameworthy loss of self-control

I intend to rule out all nine possibilities, one by one. This will reveal two
things: (1) we cannot understand the provocation defense (as theorists
typically do) in terms of the established categories of justification,
excuse, and blame; (2) we need to make room for a new category: the
morally understandable.

Before we assess each of the nine possibilities listed above, we
should consider more generally what it means to apply the concepts of
justification, excuse, and blame to a person’s mental states—specifically,
to her anger or loss of self-control. (We are better accustomed to
applying these concepts to a person’s conduct.) A person’s anger is
justifiable if it is desirable or called for, if it is morally appropriate under
the circumstances. Paradigmatically, we experience justified anger when
we observe injustice. A person’s unjustified anger is excusable or
blameless if the person can hardly be faulted for getting angry under the
circumstances even though her anger is not morally appropriate. We
experience blameless anger when we become annoyed at an
uncooperative two-year-old or at an inconvenient downpour (provided
we do not become too annoyed). A person’s unjustified anger is
blameworthy if it is something for which she can rightly be faulted, if it
evinces a bad character, if it shows her to be a bad or malicious or hateful
person—at least to some extent. We experience blameworthy anger when
we become enraged by the success of an innocent rival.
Unlike anger, loss of self-control can be excusable or blameworthy, but it cannot be justifiable. (As Richard Holton and Stephen Shute explain, “[e]ither one thinks of [losing self-control] as something that one does, in which case it is surely not justified. Or one thinks of it as something that just happens to one, in which case talk of it being either justified or unjustified is inappropriate (as one is neither justified, nor unjustified, in sneezing).”49) A person’s loss of self-control is blameworthy if she either had the ability to control herself or lacked the ability to control herself but was not blameworthy for lacking that ability. One case in which a person is blameworthy for lacking the ability to control herself is where she is blameworthy for experiencing impulses too powerful for a person of ordinary self-restraint to resist (e.g., because she habitually and voluntarily consumed addictive substances in the past). Another case in which a person is blameworthy for lacking the ability to control herself is where she is blameworthy for possessing diminished powers of self-restraint (e.g., because she habitually succumbed to ordinary temptations and thereby eroded her willpower). A person’s loss of self-control is excusable or blameless if she lacked the ability to control herself and was not blameworthy for lacking that ability.

Turning our attention to the nine possibilities listed above, we can immediately rule out (i), (iv), and (vii), each of which supposes that the provoked killer loses his self-control justifiably. Losing one’s self-control can never be justifiable. Either losing one’s self-control is something one does or it is something that happens to one. If it is something one does, then it is something the doing of which is never justifiable. (How could it ever be desirable or appropriate to give up on self-governance?) If on the other hand losing one’s self-control is something that merely happens to one, then to call it justifiable or unjustifiable is to make a category mistake.

Let’s look at the remaining possibilities. Consider first the possibility that the provocation defense sees the defendant as someone whose blameworthy anger causes him to experience a blameworthy loss of self-control [(ix)]. This is a non-starter. Clearly, such a killer is fully morally blameworthy50 and thus ineligible for mitigation under any morally

49 Holton and Shute, Self-Control in the Modern Provocation Defence, supra note _, at 70.
50 Dressler, Provocation: Partial Justification or Partial Excuse?, supra note _, at 475 (“[W]e blame—and thus do not excuse—a person who becomes angry
defensible construal of the provocation defense. The traditional defense grants no leniency to a killer whose actions “proceed from a bad or corrupt heart.” The intensity of a killer’s anger is immaterial “unless . . . the facts and circumstances were sufficient to arouse the passions of the ordinarily reasonable man,” that is, unless the killer’s anger was not wholly blameworthy. That blameworthiness also attaches to the killer’s subsequent loss of self-control only makes the killer more blameworthy overall. Ironically, the killer is perhaps even more blameworthy still where his loss of self-control was excusable [(viii)], as a killer beset by uncontrollable, malevolent anger is the very picture of a heinous criminal.

Next, consider the possibility that the provocation defense sees the defendant as someone whose justifiable or excusable anger causes him to experience a blameworthy loss of self-control [(iii) and (vi)]. To make this scenario more concrete, imagine you are justifiably (and therefore also excusably) angry about being subjected to an ugly racial epithet and you succumb to the temptation to retaliate with lethal force. Your anger is wholly appropriate, let’s suppose; yet your subsequent loss of self-control is fully blameworthy since your anger (being ex hypothesi appropriate and desirable) is surely not so intense as to be overpowering. Are you eligible for mitigation? Surely not morally: being appropriately angry is no excuse for flying off the handle. Not legally either: “[T]hough the law condescends to human frailty, it will not indulge human ferocity. It considers man to be a rational being and requires that enough to lose his self-control and kill if his anger was inexcusable [(i.e., blameworthy)].”

52 People v. Logan, 164 P. 1121, 1122 (Cal. 1917).
53 Recall the Michigan Supreme Court’s admonition: a “provocation [cannot] be held sufficient or reasonable [simply] because . . . a state of excitement [i.e., an intense passion] has followed from it; for then, by habitual and long continued indulgence of evil passions, a bad man might acquire a claim to mitigation which would not be available to better men, and on account of that very wickedness of heart which, in itself, constitutes an aggravation both in morals and in law.” Maher v. People, 10 Mich. 212, 220 (1862). See also Commonwealth v. Flax, 200 A. 632, 637 (Pa. 1938) ("[T]hough [t]he law regards with some tolerance an unlawful act impelled by a justifiably passionate heart, [it] has not tolerance whatever for an unlawful act impelled by a malicious heart.")
he should exercise a reasonable control over his passions."\textsuperscript{54} If the law mitigated the punishment of a killer whose loss of self-control was blameworthy, "the law would present no inducement to men to try to rise to the standard of even ordinary social morality."\textsuperscript{55}

Finally, consider the remaining possibilities [(ii) and (v)], according to which the provocation defense sees the defendant as someone whose blameless anger causes him to experience an excusable loss of self-control. Such a killer is certainly eligible for moral mitigation. The problem is that he is also eligible for complete moral exculpation. A killer whose blameless anger causes him excusably (and therefore blamelessly) to lose self-control is surely a person who kills blamelessly. How could an action be anything but blameless if it arises from a blameless loss of self-control caused by a blameless emotion?

Now it cannot be the case that the provocation defense supposes the defendant to kill blamelessly. If it did, provocation would have to be a complete defense instead of a partial one. So we should not accept an interpretation of the provocation defense according to which the defendant is someone whose justifiable or excusable anger causes him to experience an excusable loss of self-control. A theorist attracted to this interpretation might respond by claiming that, though the provoked killer does in fact kill blamelessly, the reason why the provocation defense mitigates the killer’s punishment instead of completely exonerating him is that, as a general matter, the law simply does not tolerate intentional homicide. But this response goes too far. The law sometimes does tolerate intentional homicide, as when it fully exculpates a person who kills in self-defense. In response, someone might point out that killing in self-defense is justified; so what the law generally does not tolerate is intentional homicide that is unjustified. But this (qualified) response also goes too far. The law sometimes does tolerate intentional homicide that is unjustified, as when it fully exculpates a person who kills while legally insane.

\textit{(b) The Morally Understandable}

We’ve now ruled out all nine ways in which the provocation defense might be understood in terms of the established theoretical categories of justification, excuse, and blame. But let us linger over (ii) and (v). For

\textsuperscript{54} R. v. Kirkham, 11 C.P. 115, 119 (1837).
\textsuperscript{55} Keenan v. Commonwealth, 44 Pa. 55, 57 (1862).
there are further reasons—more instructive ones—to reject the idea that the provocation defense sees the defendant as someone whose justifiable or excusable anger causes him to experience an excusable loss of self-control. These reasons will help us to see why the morally understandable is not equivalent to the justifiable or excusable; indeed, these reasons will help us to see why we need the concept of the morally understandable in order to make sense of the doctrine of provocation.

One of these reasons is that the provoked killer’s anger hardly seems justifiable—notwithstanding the many commentators who urge the contrary.56 Do we really believe that there are provocations in response to which it is desirable and morally appropriate to feel overwhelming, homicidal rage? I doubt it. It is one thing to claim that outrageous provocations warrant some measure of anger or indignation. It is quite another thing to claim that the kind of anger these outrageous provocations warrant is anger of such a drastic nature and intensity—anger that is (a) directed at ending the provoker’s life and (b) intense enough to cause a person of reasonable powers of self-restraint to lose control. Far from being justified, such anger seems altogether inappropriate, both in its nature (homicidal) and in its intensity.

---

56 E.g., Samuel H. Pillsbury, Judging Evil: Rethinking the Law of Murder and Manslaughter 141-42 (2001) (“Only the offender with good reason for extreme passion deserves mitigation. . . . [W]here we judge that [the offender’s] anger was not justified at the time of action, we lack one of the . . . essential elements of mitigation. . . . Simply put, provocation should require justified anger.”); John Gardner and Timothy Macklem, Compassion Without Respect? Nine Fallacies in R. v. Smith, [2001] Crim. L.R. 623, 628 (“[W]hat is held out as justified, in the law of provocation, is not the killing, but rather the loss of temper which caused the defendant to kill. She was justified in “doing as she did,” as section 3 [of the English Homicide Act] puts it, where this means “in getting so angry that she lost self-control to the point at which she killed.”); Uma Narayan and Andrew von Hirsch, Three Conceptions of Provocation, 15 Crim. Just. Ethics 15, 19 (1996) (“When an agent has been wronged, he is properly angry: it is his sense of right and wrong that gives legitimacy to his fury. . . . [H]is anger at being wronged by the eventual victim is fully justifiable: indeed, failing so to react may be a sign of the actor’s paying insufficient regard to his own worth.”); Ashworth, The Doctrine of Provocation, supra note __, at 317 (1976) (“Provocation mitigates moral culpability to the extent that a person acted in a less-than-fully-controlled manner in circumstances in which there was reasonable justification for him to feel aggrieved at the conduct of another.” (emphasis added)).
(overwhelming). To assert that such anger is justified is to accept a moral ideal according to which the virtuous person—the person who experiences all and only those emotions that he has good reason to experience—is someone who occasionally works himself into a homicidal frenzy. I doubt very much that the law (charitably interpreted) accepts any such conception of virtue. This is the virtue of some caricature of the Homeric hero or the Viking berserker, not the virtue of the ordinarily lawful person.

I also doubt that the provoked killer’s anger is even excusable, let alone justifiable. Can we really say that the provoked killer is utterly blameless for his overwhelming, homicidal anger? (Remember, the question is whether the killer is blameless for feeling anger of this particular nature and intensity. The question is not whether the killer is blameless for feeling any anger at all. Surely some measure of anger is excusable, given the extreme circumstances.) I do not think we can really say that the provoked killer is utterly blameless for his overwhelming, homicidal anger. It is undeniable that many people suffer outrageous provocations without experiencing overwhelming, homicidal anger. It is not as though every person subjected to an outrageous provocation becomes furiously angry and the only thing that separates those people who subsequently kill their provokers from those who do not is the absence of an extraordinary capacity to restrain overwhelming, homicidal anger. Overwhelming, homicidal anger is simply not an inevitable reaction to an outrageous provocation. But then what distinguishes people who become inappropriately angry from those who do not? Setting aside emotional disorders that make inappropriate anger inescapable,57 the answer appears to be that some people simply allow themselves to become inappropriately angry—or else, through failure of self-cultivation, they allow themselves to become the type of persons who tend to become inappropriately angry. (Virtually everyone is of the latter sort, in one degree or another.) But then these people are in some measure responsible for becoming inappropriately angry, and so their anger is not fully excusable.

57 We can safely ignore the issue of whether some people have emotional disorders that make it inevitable that they will experience overwhelming, homicidal anger in response to outrageous (or even trivial) provocations. Neither the case law nor the commentary contains any suggestion that a defendant must have such an emotional disorder in order to be eligible for the provocation defense.
Neither is their anger fully blameworthy, however. On the contrary, their anger is quintessentially understandable: though it shows them to be imperfectly virtuous, it does not show them to be partly vicious. If you cling to the idea that the provoked killer’s inappropriate anger is fully excusable, this might be because you assume that the only alternative is for his anger to be fully blameworthy. But there is a middle ground. There is a familiar category of moral appraisal midway between the fully excusable and the fully blameworthy. It is the category of the morally understandable. This is the category we exploit when, while apologizing for behavior that is wrongful but not malicious, we offer not an excuse but a mere explanation, an account of our motivation that explains how our behavior flowed not so much from a bad character as from an imperfectly good one. “I’m really sorry I said all those nasty things to you. I’m just so disappointed about how my life is going. I know it’s not an excuse; it’s just an explanation. I want you to know what was going through my head.”

A telling example of anger that is morally understandable yet not fully excusable emerges from Suzanne Uniacke’s discussion of John Bayley’s account of his occasional angry outbursts towards his wife, Iris Murdoch, whose obsessive and destructive behavior towards his plants was due to Alzheimer’s disease. Bayley’s outbursts were morally inappropriate, not least because on his own admission his anger was retributive, scolding. (“I told you not to! I told you not to!”) Nevertheless, while Bayley himself makes no excuses for “acting horrible things, as if kicking a child or a lamb”, we fully understand how a morally well-disposed fulltime caregiver of someone with advanced Alzheimer’s disease could occasionally break down in anger in just the way that he did.  

Uniacke goes on to suggest (wrongly, in my view) that “we are inclined to excuse [Bayley’s anger] . . . as an understandable occasional reaction of a normal human being towards behavior that would try the patience of a saint.”  

I think Uniacke goes too far in supposing that we are inclined to find Bayley’s anger excusable. No doubt we feel great sympathy for Bayley and we hardly think his retributive anger shows him to be a malicious person. But we blame Bayley for his retributive anger, even as we pity him for being in the situation that elicited it. We regard Bayley’s

---

59 Id. at 115 (emphasis added).
anger (as he himself regards it) as something for which he is responsible, hence, as something about which he rightly feels guilty. At the same time, we recognize Bayley’s anger as something natural and difficult to avoid, as something for which he should not be faulted too severely. So to some extent we excuse him. But if we excuse him only to some extent, then to some extent we blame him. This will seem paradoxical only if we think of blameworthiness as ‘all or nothing.’ If we think of it instead as occupying a continuum, from the fully blameless (excusable) to the fully blameworthy, then we will see that there is an expansive middle range: the domain of the morally understandable.

The morally understandable is a category we would do well to emphasize in legal theory, for it is already present in the law. It is what courts are after when they describe an “adequate provocation” as one sufficient to arouse the ire of an “ordinary” person, if not necessarily a “reasonable” one.60 And it is what Lord Millett was after when he offered the following account of the provocation defense’s moral structure:

[It] can never be reasonable to react to provocation by killing the person responsible. Nor by pleading provocation does the accused claim to have acted reasonably. His case is that he acted unreasonably but only because he was provoked. But while this may not be reasonable it may be understandable, for even normally reasonable people may lose their self-control and react unreasonably if sufficiently provoked.61

As Lord Millett here implies—and as I have already indicated—the provocation defense applies the ‘understandable-ness’ test not just to the provoked killer’s anger but also to his subsequent loss of self-control. That is, the provocation defense envisions the killer’s loss of control as merely understandable—not fully excusable.

---

60 E.g., State v. Gounagias, 135 P. 9, 12 (Wash. 1915) (“[T]he question of adequate or reasonable cause [i.e., provocation] is essentially a question of fact, to be submitted to the jury under proper instructions whenever it can be said that the alleged provocation would have any reasonable tendency to produce sudden and uncontrollable anger and heat of blood in the ordinary man.”); Maher v. People, 10 Mich. 212, 216 (1862) (“[A]dequate provocation . . . [is] anything the natural tendency of which would be to produce [intense anger] in ordinary men. . . .”).

The reason why I deny that the provoked killer’s loss of control is fully excusable is not the reason that commentators typically give for denying this, namely, that if the killer’s loss of control were fully excusable, then the provocation defense would have to be fully exculpatory. That just isn’t the case. True, if the killer’s loss of control were fully excusable, then his homicidal impulse would be irresistible and his loss of control, unavoidable. But these facts alone would not entitle the killer to a full excuse. To be entitled to a full excuse, the killer would not only have to suffer a fully excusable loss of self-control, he would also have to act from fully excusable anger; for, as we saw above, if the killer’s anger isn’t itself fully excusable, then his subsequent loss of control won’t be exculpatory—even if that loss of control is unavoidable. Indeed, if the killer’s blameworthy anger is so intense that his subsequent loss of control is unavoidable (hence, itself fully excusable), that fact may serve to aggravate his guilt rather than to mitigate it.

So the reason why the provocation defense envisions the provoked killer’s volitional impairment as merely understandable is not, then, that blameless volitional impairment would militate in favor of making the provocation defense fully exculpatory. As we just saw, it wouldn’t. Rather, the reason why the provocation defense envisions the provoked

---

62 Contra Dressler, Why Keep the Provocation Defense?, supra note __, at 974 (“[T]he [defendant’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. If he were totally incapable, a full excuse would be defensible.”); Kent Greenawalt, “Uncontrollable” Actions and the Eighth Amendment: Implications of Powell v. Texas, 69 Colum. L. Rev. 927, 962 n.171 (1969) (“Even if provocation that overpowered the will were a complete defense [i.e., a defense to crimes beside murder], mitigation, rather than exoneration, would be appropriate if provocation made an impulse difficult, but not impossible to resist.”); Garvey, Passion’s Puzzle, supra note __, at 1702 (arguing against the so-called “total-incapacity theory” of the provocation defense).

63 See supra note ___ and accompanying text. See also Maher, 10 Mich. at 217 (“[I]t is not the case that the provocation, in every case, [must] be held sufficient or reasonable, [just] because . . . a state of excitement has followed from it; for then, by habitual and long continued indulgence of evil passions, a bad man might acquire a claim to mitigation which would not be available to better men, and on account of that very wickedness of heart which, in itself, constitutes an aggravation both in morals and in law.”).
killer’s volitional impairment as merely understandable is the simple but vital fact that his homicidal impulse is not likely to be truly irresistible.\textsuperscript{64} Consider the reflections of psychologist Roy F. Baumeister, who has spent decades studying self-control:

[In research on self-control, one conclusion stands out over and over again: People acquiesce in losing control. In other words, they let themselves lose control, and they become active participants. Whether it is a matter of breaking a diet, going on a drinking binge, or abandoning an unpleasant task, usually the person somehow allows it to happen. The same applies to violence. The concept of an irresistible impulse is somewhat misleading, because most violent behavior is not truly the result of irresistible impulses. People allow themselves to lose control. And they do so in part because they want to regard certain impulses as irresistible.

Often the person’s acquiescence in losing control is cleverly disguised. For example, people speak as if an eating or drinking binge were simply a matter of being overwhelmed by strong impulses that rendered them passive and helpless. Yet during these binges, they continue to procure food or drink, prepare it for consumption, put it in their mouths, and swallow it. These are active, not passive actions. Resisting the impulse may have been too difficult for them, but they have not simply quit resisting, they have become active accomplices in indulging their desires.

The same is often true of violence. Many people speak as if certain provocations produce an unstoppable rage that make[s] violent action impossible to stop. . . . Most of the time, they do actually retain some degree of control. For example, an angry husband may say that he was so provoked by his wife that he could not stop himself from beating her severely, but in fact he does usually stop himself before he permanently injures or kills her. Whatever his heat of passion, he usually does know when to stop.\textsuperscript{65}


\textsuperscript{65} Baumeister, Evil, supra note _, at 274-75.
Now if the provoked killer’s violent impulse isn’t truly irresistible, then his volitional impairment can’t be truly blameless: if he should have resisted his violent impulse—and could have resisted it—then his loss of control is not fully excusable. But neither is it fully blameworthy. The violent impulse is insistent and powerful—so powerful as to “render ordinary men of average disposition liable to act rashly or without due deliberation and reflection, and from . . . passion rather than from judgment.”66 That someone would acquiesce to such an impulse is surely understandable, surely less blameworthy than if the impulse were not so insistent and so powerful.

We might compare the provoked killer to someone in the grip of a potent addictive desire. Though addictive desires are generally not irresistible—some addicts can abstain for days or even weeks when their circumstances demand it—these impulses are incredibly powerful, certainly powerful enough to “render ordinary men of average disposition liable to act rashly or without due deliberation and reflection, and from . . . passion rather than from judgment.”67 For this reason, when a person acquiesces to an addictive desire, we feel pity and compassion, even if we do not find him altogether blameless. We regard the addict’s volitional impairment as sitting somewhere between the fully excusable and the fully blameworthy. Indeed, we often blame the addict for losing control even where we do not fully blame him for being an addict—even where, say, his addiction resulted from taking duly-prescribed painkillers. Thus, even where the addict is not fully to blame for the fact that he has trouble exercising self-control, his eventual failure to control himself can still be somewhat blameworthy. Likewise for the provoked killer: even though he is not altogether blameworthy for the fact that he has so much trouble exercising self-control—he was provoked, after all—neither is he altogether blameless for failing to control himself.

At its moral core, the provocation defense sees the provoked killer as being less than fully blameworthy where his anger and volitional impairment are at once partly blameworthy and partly excusable—where they are morally understandably. Hence, the idea that the provocation defense involves “a concession to the frailty of man, a recognition that

---

66 People v. Logan, 164 P. 1121, 1123 (Cal. 1917).
67 Id.
the average person can understandably react violently to a sufficient wrong...”

2. The Provocation Requirement [(6)]. — If both a defendant’s anger and his volitional impairment are morally understandable, these facts suffice to mitigate his moral guilt. But they do not suffice to mitigate his legal guilt. He must also satisfy the provocation requirement. No matter how understandable are a person’s anger and loss of self-control, he cannot avail himself of the provocation defense unless his anger was aroused by conduct that constituted a legally cognizable provocation.

The early common-law authorities apparently recognized only four categories of “adequate provocation”:

The first category was the general one of a grossly insulting assault.
The second category can broadly be described as seeing a friend, relative, or kinsman being attacked. The third category was . . . seeing an Englishman unlawfully deprived of his liberty. The fourth category was seeing a man in the act of adultery with one’s wife.69

Modern courts take a more open-ended approach: instead of ruling just a few categories in, they rule just a few categories out. Among the categories excluded are: “words alone, however scurrilous or insulting,”70 a bare trespass to the defendant’s property,71 a provocation created or incited by the defendant,72 and (in some jurisdictions) a romantic or sexual advance on the defendant by a person of the same sex.73 Pursuant to the three-part classical gloss, modern courts rationalize these categorical exclusions on the grounds that the excluded “provocations” are not “adequate,” where this means either that intense, homicidal anger is never a reasonable emotional reaction to the excluded “provocations” (the evaluative interpretation) or that the excluded “provocations” could never cause a person endowed with ordinary powers of self-restraint to have serious difficulty controlling herself (the volitional interpretation).

68 State v. Guido, 191 A.2d 45, 55 (N.J. 1963); see also Maher, 10 Mich. at 212 (describing the provocation defense as an “indulgence to the frailty of human nature”).
69 Horder, Provocation, supra note _, 24.
70 Sells v. State, 653 P.2d 162, 164 (N.M. 1982).
72 See Roberson v. State, 117 So. 412, 414 (Ala. 1928).
73 Kahan and Nussbaum, Two Conceptions of Emotion, supra note _, at 310.
Shifting to the idiom of the six-part formula, we might accordingly be tempted to rationalize the categorical exclusions on the grounds that the excluded things could never arouse homicidal rage that was morally understandable. But can we really say that? Can we really say that it is never morally understandable for a person to become severely, homicidally angry in response to mere words—even where the words are a vicious anti-Semitic tirade and their target has just escaped a Nazi death camp? I don’t think we can. Nor do I think we must. Not only is categorical exclusion something that the emotional requirement can’t comfortably pull off, it is also something that the emotional requirement doesn’t need to pull off. Categorical exclusion is better done by the provocation requirement, which (unlike the classical element of “adequate provocation”) is not in any respect the emotional requirement in disguise.

What’s at stake here is not merely theoretical. One practical advantage of delegating exclusionary work to the provocation requirement is that the provocation requirement presents a question of law for the judge. As such, the provocation requirement empowers the judge to circumvent a jury’s bias (or worse), where this bias might incline the jury to conclude that a defendant’s homophobic or racist or misogynistic fury was morally understandable. No doubt the judge might not be any less homophobic or racist or misogynistic than the jury; yet the judge’s ruling will be subject not only to the prevailing law on what counts as a legally cognizable provocation but also to public criticism of a sort unlikely to have much effect on a jury but potentially career-ending for a judge.74

II. PROVOCATION’S PSYCHOLOGICAL FOUNDATION

In the previous Part, I offered a doctrinal and moral defense of a novel version of the provocation defense, the six-part formula:

In a prosecution for murder, it is a defense that the defendant
killed as a result of
(1) a loss of self-control caused by
(2) anger aroused by
(3) the victim’s conduct,
provided that, as determined by the finder of fact,

(4) the defendant’s loss of self-control was morally understandable given the nature and intensity of her anger and
(5) the nature and intensity of the defendant’s anger were morally understandable given the victim’s conduct, and,
provided that, as determined by the finder of law,
(6) the victim’s conduct constituted a legally cognizable provocation.

As I argued, the six-part formula reveals the doctrine’s full psychological complexity, and it does so in a way that makes the defense’s moral structure transparent. At its moral core, the provocation defense deems certain killers less than fully blameworthy if their homicidal conduct resulted from morally understandable volitional impairment caused by morally understandable anger.

To put the doctrine on a firm psychological footing, we must articulate a theory of the provoked defendant’s retaliatory anger that makes it intelligible how such anger could both (i) be morally understandable in light of the victim’s prior conduct [the emotional requirement] and (ii) be the cause of morally understandable loss of self-control [the volitional requirement]. This is not as easy as it might seem. As I will show, the main philosophical theories of emotion and desire—the main theories of temptation—do not in fact allow for the possibility of morally understandable loss of self-control. The leading theory of temptation (the evaluative conception) does not allow for volitional impairment at all, while the main rival (the mechanistic conception) allows only for volitional impairment that is fully excusable.

A. The Evaluative Conception of Temptation

The evaluative conception is the dominant approach to understanding the emotions (in empirical psychology as well as in philosophy) and it is increasingly influential as a philosophical theory of desire. For evaluative theories of emotion, see, e.g., Plato, The Republic; Aristotle, On the Soul; 2 The Hellenistic Philosophers 404-18 (A.A. Long & D.N. Sedley eds., 1987) (on the Stoics); C.D. Broad, Emotion and Sentiment, in Broad, Critical Essays in Moral Theory (1971); William Lyons, Emotion (1980); Robert C. Solomon, The Passions: Emotions and the Meaning of Life (1993); Jerome Neu, A Tear is an Intellectual Thing (2000); Christine Tappolet, Emotions et Valeurs (2000); Martha C. Nussbaum, Upheavals of Thought: The Intelligence of Emotions (2001); Jesse J. Prinz, Gut Reactions: A Perceptual
The Provocation Defense

Mendlow

criminal law. Kahan and Nussbaum see the evaluative conception as undergirding not just the provocation defense but also the defenses of duress, insanity, and self-defense.\textsuperscript{76} As Kahan and Nussbaum explain the evaluative conception of emotion,

> the emotions themselves contain an evaluation or appraisal of [their objects]—that is, the appraisal is part of the belief in terms of which the emotion will be defined, and these ways of seeing the world are part of what the emotional experience includes. Grief sees the lost one as of enormous significance; so too, in a happier way, does love. Disgust usually sees the object as one that threatens or contaminates, one that needs to be kept at a distance from the self. Fear perceives the impending harm as significant; anger sees the wrong as pretty large—whether or not this is the way these things really are.\textsuperscript{77}

In a similar vein, the evaluative view of desire conceives a desire to be a quasi-cognitive (or perhaps quasi-perceptual) way of regarding the desired object as good or valuable. According to this view, to desire something is to see it as good or valuable, or to see it as something that

\textsuperscript{76} See, e.g., Kahan and Nussbaum, Two Conceptions of Emotion, supra note _, at 285.

\textsuperscript{77} Kahan and Nussbaum, supra note _, at 285.

there is reason to obtain or bring about. Common to the evaluative views of desire and emotion is the idea that temptation’s essence is evaluative appearance: temptation makes its object appear good or look good or seem good. Whether we really believe that the object is good—this is a further question. An object can appear good without our really believing that it is.

The evaluative conception has little trouble explaining the provocation defense’s emotional requirement, which demands that the provoked defendant’s anger be morally understandable in light of the victim’s conduct. According to the evaluative conception, the provoked defendant’s anger is morally understandable because it embodies both the wholly correct judgment that the victim’s conduct was wrongful and the partly correct judgment that the victim deserves some measure of adverse treatment. The latter judgment is only partly correct because the adverse treatment warranted by the victim’s conduct is legal punishment, not extrajudicial execution.

While the evaluative conception handily explains the emotional requirement, it has much more trouble with the volitional requirement. The reason for this is that the evaluative conception cannot explain how anger qua evaluative appearance could lead us to act akratically, that is, voluntarily against our better judgment. Or so I will argue.

I should pause for a moment to explain why the prototypical case of provoked retaliation is a case of akrasia, a case of voluntary action contrary to one’s better judgment. Rarely does the provoked defendant suddenly change her mind about whether homicide is morally and legally wrong; ordinarily she remains cognizant of homicide’s wrongness, even if other cognitions become far more salient. The main reason why we might think otherwise is that the provoked defendant typically believes that her provoker deserves to be harmed or killed. Here we

---

78 Though philosophers have occasionally offered arguments to show that akrasia is impossible, see Plato, Protagoras 358b-c; R.M. Hare, The Language of Morals (1952); R.M. Hare, Freedom and Reason (1963), none of these arguments commands anything close to widespread assent. Moreover, no empirical evidence of which I am aware indicates that people never act akratically.

79 I disagree with Stephen P. Garvey, Passion’s Puzzle, 90 Iowa L. Rev. 1677, 1683 (2005) (asserting that the provoked defendant is often someone who “violates the law … because he honestly and momentarily … believes the law allows him to kill”).

34
would do well simply to bear in mind that a person can believe fervently and passionately that her provoker deserves to be killed while at the same time believing that it would be wrong for her to be the one to kill him. (Consider the difference between (i) believing that executions are sometimes justified and (ii) believing that you have the right to be an extrajudicial executioner.) Even in moments of intense anger, we rarely stop believing that it is wrong to take the law into our own hands. But even when we do stop believing this—even when we do temporarily change our minds about the moral permissibility of revenge killing—we do not change our minds about whether it would be a good idea to perform an action that will very likely send us to prison for a long, long time. The point is this: even if you suppose that the provoked defendant suddenly and temporarily forms the belief that revenge killing is morally permissible, it strains credulity to suppose that she also forms the belief that it is worth going to prison for many years in order to perpetrate this particular revenge killing.\(^80\)

Proponents of the evaluative conception tell a standard story about how a temptation qua evaluative appearance leads us to act akratically. The standard story draws an analogy between akratic action and perceptual illusion.\(^81\) As Christine Tappolet explains, akratic action involves “a conflict between a value perception and an evaluative judgment that can be compared to perceptual illusions such as the Muller-Lyer illusion, in which one sees the lines as being of a different length.”

---

\(^80\) Cf. Holton and Shute, Self-Control and the Modern Provocation Defence, supra note _, at 60 n.39 (“We . . . need to distinguish cases of [loss of self-control] from cases in which the agent simply judges that the best course is to respond to provocation with violence, and does so. . . . [In such cases], the provocation defence cannot be invoked . . . for the simple reason that there is no loss of control.”).

\(^81\) Versions of this story appear in: Christine Tappolet, Emotions and the Intelligibility of Akratic Action, in Stroud and Tappolet, supra note _, at 111; Scanlon, What We Owe, supra note __, at 35; Wallace, Addiction, supra note __, at 185-86; Johnston, supra note __, at 214; Tenenbaum, Accidie, supra note __; Talbot Brewer, The Character of Temptation: Towards a More Plausible Kantian Moral Psychology, 83 Pac. Phil. Quarterly 103 (2002). See also Jessica Moss, Akrasia and Perceptual Illusion, 91 Archiv f. Gesch. d. Philosophie 119, 119 (2009) (arguing that Aristotle’s account of akrasia entails that “akrasia is parallel to certain cases of perceptual illusion.”).
length even though one judges or even knows that they are of the same length. Tappolet illustrates this comparison with several examples:

Suppose I am about to cross a narrow rope bridge hanging high up on a deep shaft. Though I feel fear, I judge that all things considered I ought to cross the bridge; I judge it to be sufficiently safe and going back would make for a much longer hike. If I end up not crossing the bridge, it will not be difficult to make sense of my action: the perceived danger, be it real or not, readily explains why I don’t cross the bridge. Or consider . . . [Dante’s] Francesca and her passionate love for Paolo . . . [T]his love consists in the perception of Paolo as a worthy object of love. Now this cannot directly make Francesca’s action intelligible, for it does not involve a perception of the value of the action itself. But it is surely an important part of what makes it intelligible. The value that Francesca perceives makes her desire to make love to Paolo intelligible and thus indirectly makes her action intelligible . . . even though [she] judge[s] that another course of action would have been better all things considered.

These examples are meant to show that temptation qua evaluative appearance causes us to act akratically by functioning as a sort of practical illusion—an illusion that presents certain considerations as being weightier than we believe them to be. If the provoked defendant acts akratically, then his retaliatory desire presents him with an illusion whose content is that retaliation is justified. He knows that retaliation is in fact unjustified or at least a very bad idea—he knows, in other words, that the illusion is an illusion—but he retaliate anyway, because the practical illusion is especially vivid.

The standard story has a certain theoretical elegance. Yet it ultimately fails to explain akrasia because it fails to make sense of how we could be led astray by evaluative illusions that we recognize as illusions. If I really do believe that Y is better than X—as I must, if my pursuit of X is to be truly akratic—how can X’s seeming or appearing good prompt me all by itself to pursue X? As we saw, Tappolet and others answer this question by exploiting a supposed analogy between, on the one hand, evaluative appearance and action and, on the other, perceptual illusion and belief. But this strategy suffers from a rather serious problem: we would quite likely reject the counterpart story about perceptual illusion and belief.

---

82 Tappolet, supra note _, at 111.
83 Tappolet, supra note _, at 111.
To see why, suppose that I encounter a white marble statue that is bathed in green light in such a way as to make the statue appear green—a workaday perceptual illusion. Suppose further that I do not believe that the statue really is green. I have seen the statue under normal lighting conditions and I know that no one has tampered with the statue (for instance, by painting it green). Despite this, could the statue’s (merely) appearing green prompt me to change my mind about what color the statue really is?

I do not see how it could. Absent other factors—a ‘credulousness’ drug, reasonable doubts about whether the statue was modified before I entered the room, and so forth—I do not see how the perceptual illusion could prompt me to revise my belief. For that matter, I do not see how the perceptual illusion could even tempt me to revise my belief.

The point is not that perceptual illusions always fail to make us revise our beliefs. They succeed in this quite often, of course. But never when we recognize them as illusions. When I see my reflection in a mirror, I am not tempted to believe that my doppelgänger is staring at me. Even illusions that are vivid and terrifying cannot make us revise our beliefs if we know them to be illusions. I once went on a ride at Disney Land that used holographs to give riders the sensation that they were hurdling toward a head-on collision with a speeding train. Though I ducked and flinched and screamed in fear, I did not believe one bit that the train was really there. Nor was I even tempted to believe it. My behavioral and visceral responses to the holographic image of the train were hardly manifestations of belief. Few if any of the cognitive hallmarks of belief were present. I was not disposed to assert that the train was really there, for example, nor did I treat the proposition that the train was really there as a premise in further reasoning. The holographic image of the train only seemed to be a reason to believe that the train was hurdling toward me, and so I did not form the belief that it really was.

These reflections suggest that the standard story, with its analogy between akratic action and illusion, is not just unhelpful to the evaluative conception; it is fatal. Far from showing how akratic action is intelligible, the standard story appears to entail that akratic action is impossible. For if a known perceptual illusion cannot tempt or cause belief, then—taking seriously the analogy between evaluative appearance and perceptual illusion—it would seem to follow that a known illusory evaluative appearance cannot tempt or cause action. In particular, it would seem to follow that, no matter how good or right or justified some
prospective act of violent retaliation might appear, this illusory appearance cannot tempt me to retaliate. The basic point is this: a thing’s seeming good (when I know it is not) can no more tempt me to perform an action than a thing’s seeming green (when I know it is not) can tempt me to form a belief.

“But perceptual illusion really can tempt belief,” an evaluative theorist might insist. “When I see a straight stick that is partially submerged in a bucket of water, I am ever so tempted to believe the stick is bent, even though I know very well it is not.” Then the evaluative theorist and I are quite different, I will reply. Yet it may be I who am the eccentric one, so I will grant for the sake of argument that perceptual illusion can tempt belief. Still, this does nothing to show that any perceptual illusion, however tempting, can actually cause belief. No matter how tempted you are, surely you will not come to believe that the stick is really bent. Surely nothing could be easier than resisting this doxastic (i.e., belief-related) temptation. But if nothing could be easier than resisting this doxastic temptation, then—taking seriously the analogy between evaluative appearance and perceptual illusion—it would seem to follow that nothing could be easier than resisting a practical temptation. It would seem to follow, in particular, that nothing could be easier than ignoring an ‘illusion’ to the effect that violent retaliation is justified or right. But if it were true that nothing could be easier than ignoring a practical illusion, this would make resisting temptation so incredibly easy that we would find it utterly mysterious—indeed unintelligible—how anyone could ever act akratically.84

Because it renders akrasia impossible or at least unintelligible, the evaluative conception evidently cannot undergird the defense of provocation, which (I submit) presumes that the prototypical act of retaliation is both akratic and intelligible. What could be more intelligible than an act of vengeful retaliation? What could make more sense than this—both from the standpoint of the actor and from the perspective of an onlooker?

B. The Mechanistic Conception of Temptation

The evaluative conception founders on its inability to explain volitional impairment. The mechanistic conception promises to do better

---

84 This might help explain why evaluative theorists from Aristotle to Donald Davidson have found akrasia so puzzling.
because it identifies temptation with something quite similar to volition itself. As we will see, the mechanistic conception does explain volitional impairment. Yet the mechanistic conception makes temptation difficult to resist only at the cost of making it impossible to resist. This conflicts with the volitional requirement, which adheres to the idea that the provoked defendant’s loss of control is merely understandable rather than fully excusable.

The mechanistic conception shows up in the work of Descartes, Kant, and Freud, as well as in the behaviorist and neo-behaviorist psychology of the twentieth century. Its most ardent contemporary defender is probably the philosopher Harry Frankfurt, who explains the mechanistic conception like this:

However imposing or intense the motivational power that the passions mobilize may be, the passions have no inherent motivational authority. Considered strictly in themselves, apart from whatever additional impetus or facilitation we ourselves may provide by acceding to them, their effectiveness in moving us is entirely a matter of sheer brute force. There is nothing in them other than the magnitude of this force that requires us, or that even encourages us, to act as they command.

Elsewhere Frankfurt advances a mechanistic conception specifically of “our most elementary desires”:

Animals of many species have desires, but only animals of our species—or, perhaps, of a few others—are capable of seeing anything as a reason. Our most elementary desires come to us as urges or impulses; we are moved by them, but they do not as such affect our thinking at all. They are merely psychic raw material. A desire provides us not with a reason but with a problem—the problem of how to respond to it. Impulses and urges have power, but in themselves they have no authority. They move us more or less strongly, but they make no claims on us.

---

85 See, e.g., René Descartes, Passions of the Soul; Sigmund Freud, A General Introduction to Psycho-Analysis (1935). See also Brewer, supra note _ (discussing Kant’s mechanistic account of desires). See generally Kahan and Nussbaum, supra note _, at 280-82 (offering a concise history of the mechanistic conception of emotion, replete with references to contemporaries and historical figures).
87 Harry G. Frankfurt, Reply to T. M. Scanlon, in Buss and Overton, supra note _, at 184.
So conceived, temptations consist solely in felt motivation: “They are merely psychic raw material” whose “effectiveness in moving us is entirely a matter of sheer brute force.” For this reason, temptations mechanistically conceived have no normative authority: they do not in themselves give rise to practical reasons or requirements. They in fact lack even the mere pretence of normative authority; they do not even seem to give rise to practical reasons or requirements. Nor do they present their objects as good or valuable. Indeed, not only do temptations mechanistically conceived fail to present their objects as good or valuable, they fail to present their objects even as appealing or attractive: “There is nothing in them other than the magnitude of [their sheer brute] force . . . that even encourages us . . . to act as they command.”

How might such vectors of sheer brute force lead us to act akratically? One possibility—which in the end I will argue is the only true possibility—is that vectors of sheer brute force lead us to act akratically by being overpowering compulsions, impulses that are literally irresistible. This possibility is of course a non-starter for anyone who would care to remain faithful to the idea that no matter how difficult resisting temptation sometimes is, it is not ordinarily impossible. (As I urged above, the temptation to retaliate against an outrageous provocation is not ordinarily irresistible.)

Consider instead the possibility that temptation qua vector of sheer brute force leads us astray not through compulsion but through what we might call ‘diversion.’ Just as an extra-personal force like a gust of wind can be strong enough to dissuade me from forging ahead without being so strong as to blow me off my feet, so might an intra-personal force be strong enough to divert me from my course without wholly compelling me to give it up. If this is the mechanism by which temptation operates, then I act on a temptation (when I do) not because I can’t resist it—I can—but because I don’t. And why don’t I? Because I choose not to. Because I acquiesce. So when I strike back against an outrageous provocation, I strike back not because my retaliatory desire overpowers me but because I yield to its power.

This ‘diversion’ model is an obvious improvement on the ‘compulsion’ model because it does not depict temptations as overpowering compulsions. Yet the ‘diversion’ model cannot but render akratic action less than full-blooded—not because it renders akratic action involuntary but because it renders such action mysterious, bizarre, or unintelligible. If a temptation qua vector of sheer brute force is
resistible, then why on earth do I not resist it? You will strain to make sense of my acquiescence and, crucially, so will I. Conceived as a vector of sheer brute force, a temptation does not entice or seduce, does not present its object as appealing or attractive. Much less does it present its object as good or valuable. That I would act on such a temptation in defiance of my all-things-considered judgment is bizarre. It is something I will not be able to make any sense of—something I will find unintelligible. The problem here is not the problem of perversity, the problem of how a person could intelligibly pursue the bad for the sake of nothing but its badness. That is the problem of how a person could intelligibly find the bad appealing. The present problem is worse. It is the problem of how a person not acting on an irresistible compulsion could in defiance of his all-things-considered judgment intelligibly pursue what he does not find appealing at all. If acquiescence to such an urge is unintelligible in this sense, it is hardly morally understandable.

To be clear, I am not rehearsing the well-worn charge that “a mere behavioural disposition [cannot] make the behaviour it causes intelligible.” This is a charge that philosophers customarily substantiate by appealing to examples like that of Warren Quinn’s ‘radio man,’ a fantastical character beset by an inexplicable, affectless urge to turn on every radio he encounters. For my part, I do not share the intuition that affectless urges are by their nature apt to cause behavior that is unintelligible. On the contrary, I find it rather obvious that people often act intelligibly on brute urges, urges that do not depict their objects as appealing or attractive. People hum tunelessly under their breath, drum their fingers on the table, avoid stepping on cracks in the sidewalk—all because of brute urges. None of this is unintelligible or even out of the ordinary. But what makes this behavior intelligible and ordinary is that it is harmless. None of it contravenes the actor’s all-things-considered practical judgment. If, on the other hand, a person thinks it harmful to waste his energy on avoiding the cracks in the sidewalk yet perseveres in this avoidance solely because of a resistible brute urge, his behavior will surely be unintelligible—not just to others but to himself as well. For it is utterly mysterious how a person not acting on an irresistible compulsion

---

88 Tappolet, supra note __, at 112. On this point, Tappolet cites: Warren Quinn, Putting rationality in its place, in Quinn, Morality and Action (1993), at 236-37; Scanlon, What We Owe, supra note __, ch. 1; Jonathan Dancy, Practical Reality (2000), ch. 2; Johnston, supra note __.
89 Quinn, supra note __, at 236-37.
could in defiance of his all-things-considered judgment intelligibly pursue what he does not find appealing at all. The only thing that could make this behavior intelligible, it seems, would be for the temptation to be irresistible; only if the impulse were irresistible would it make any sense for the person to acquiesce. But if a person can act contrary to his better judgment only when he is afflicted by an irresistible impulse, then it hardly seems that he can act contrary to his better judgment voluntarily. And if he cannot act contrary to his better judgment voluntarily, then he cannot truly act akratically. For without the element of voluntariness, *akrasia* collapses into compulsion, hence, loss of control seems fully excusable.

A mechanistic theorist might object to all this as follows: “You have assumed that resisting a temptation is always costless. But your assumption is false. Far from costless, resisting a temptation may be difficult or taxing or painful. We may choose not to resist, then, simply because we don’t want to go through the trouble. What makes such a choice akratic is that, by our own lights, the difficulty of resistance is not great enough to justify our acquiescence. We acquiesce not because we deem resistance unjustifiably difficult—we don’t—but because we are lazy or fearful or selfish. This choice may be unwise or immoral (although it need not be). But it surely isn’t unintelligible.”

This is a reasonable point, but no mechanistic theorist can make it. The reason is that the mechanistic conception applies no less to a person’s aversion to the discomfort of resisting a given temptation than it does to that very temptation. The aversion is itself just another temptation, so the aversion, too, must be a (resistible) brute urge. But if the aversion is itself a resistible brute urge, then the problem of unintelligibility will simply reassert itself. If we cannot intelligibly acquiesce to the brute urge that is the underlying temptation, neither can we intelligibly acquiesce to the brute urge that is our aversion to resisting the underlying temptation. Only if the urge were irresistible could we acquiesce to it intelligibly.

C. Psychic Attraction and the Provocation Defense

To avoid the problems that beset the evaluative and mechanistic conceptions, we need a theory of desire that allows for *affect* (unlike the mechanistic conception) but does not reduce affect to *appraisal* (unlike the evaluative conception). I offer such a theory here.
I. A Theory of Affective Desire. — Philosophers have at one time or another attached the label ‘desire’ to just about every kind of motivating state. My target is of course much narrower. As Mark Johnston observes, there is a perfectly good non-philosophical sense of ‘desire’ in which desire is not only one of the springs of action, but a state which makes certain kinds of actions readily intelligible. In this sense of ‘desire’, which we might distinguish by the somewhat pleonastic name ‘affective desire’, we desire other things and other people, we are struck by their appeal, we are taken with them.90

Desire in this sense is a thing we fight, indulge, and condemn; a thing that makes our lives colorful, meaningful, and difficult. Though there are other kinds of motivating state that philosophers have sometimes called by the name ‘desire’—those arising from our decisions, intentions, and practical judgments, for example91—these other motivating states fall outside the scope of my inquiry. The urge to retaliate against an outrageous provocation is of course a prime exemplar of affective desire. As I develop the general theory, however, I will say nothing about the retaliatory urge, since I do not want to be seen as stacking the deck. Only after my theory of affective desire is fully in view will I apply it to the case of provocation.

My chief contention is that the essence of affective desire (hereafter, simply ‘desire’) is a feeling of psychic attraction. As I will argue, this feeling is not reducible to any combination of the following: a feeling of motivation; an appearance of the desired object as good or valuable; an ascription of appealing properties to the desired object; or a state of insistent attention to the object’s appealing properties. These phenomena ordinarily accompany desire, but none is the thing itself. I suggest that the feeling of psychic attraction is itself indefinable, that it can’t be analyzed in terms of more primitive psychic phenomena. Here I take a page from David Hume, who made a similar observation while discussing pride and humility:

The passions of PRIDE and HUMILITY being simple and uniform impressions, ‘tis impossible we can ever, by a multitude of words, give a just definition of them, or indeed of any of the passions. The utmost

90 Johnston, supra note _, at 188.
we can pretend to is a description of them, by an enumeration of such
circumstances, as attend them…”

Something similar seems true of desire more generally: though we may
not be able to define it, we can describe the phenomena that ordinarily
attend or characterize it. In this respect, desire is no different from a host
of familiar psychological states. The emotions each have a distinctive but
ineffable feel and so do the many varieties of pain and pleasure. Though
there is much we can say about the cognitive and volitional conditions
that characteristically attend these psychological states, there is little we
can say about their felt quality—little that is not hopelessly metaphorical.
So instead of describing psychic attraction, I will isolate it—
distinguishing it from the phenomena by which it is typically
accompanied and for which it is often confused.

We can make some headway in isolating psychic attraction if we
scrutinize a single case. Consider thirst. In adult humans, thirst is a
syndrome that involves at least an experiential state (an unpleasant
sensation of dryness in one’s mouth and throat) and a thought (the
thought that a certain activity would relieve the unpleasant sensation).
Crucially, thirst also involves a feeling of attraction—whether to
drinking (an act), to potable liquid (a thing), or to one’s thirst being
slaked (a state of affairs). For our purposes, it doesn’t much matter how
we conceive of thirst’s intentional object. What matters is how we
understand the feeling of attraction.

A natural thought is that we should understand the feeling of
attraction as a kind of felt motivation. This thought derives some initial
plausibility from the obvious fact that a thirsty person’s desire for drink
grounds a set of motivational dispositions, dispositions to seek out
potable liquids, for example, and to consume them. But it is a further
question whether a thirsty person’s felt attraction to drinking is nothing
but the conscious manifestation of these motivational dispositions. And
the answer appears to be no. Felt motivation seems insufficient for
psychic attraction, as I might feel myself motivated to pursue things to
which I am not the least bit attracted. Suppose that, suffering from a
disease like pica, I am afflicted by a brute urge to consume certain vile-

---

92 David Hume, A Treatise of Human Nature (L.A. Selby-Bigge and P.H.
Nidditch eds., 1978), at 277.
93 Scanlon, What We Owe, supra note _.
tasting fluids, such as glue and liquid laundry detergent. Though I find these fluids disgusting, I am nevertheless beset with a resistible but insistent impulse to drink them. My urge would seem hardly to be a case of desire, since there is no respect in which I find the prospect of drinking appealing. This is by no means to suggest that brute urges are always bizarre or pathological. I may have a brute urge to fidget with my wedding ring or to click my pen repeatedly. When I do these things, I do them simply because I feel so motivated, not because the actions strike me as appealing. Now the impulses that prompt these actions are surely not pathological. They are normal. But that doesn’t make them desires. When I desire something, I am not merely motivated to pursue it. I am taken with it. I find it appealing.

What is it, then, to find something appealing, to be psychically attracted to it? If finding something appealing is not a matter of feeling a motivation to pursue the thing, is it instead a matter of the thing seeming good or valuable? It is certainly natural to say of the things we desire that they “seem good.” When we desire something, it almost always strikes us as good in some respect and to some extent, even if we simultaneously judge that the thing is not actually good at all. Indeed, it might well be true that we can desire something intelligibly only if we find it good in some way or other. But if being psychically attracted to something were simply a matter of finding it good, psychic attraction could not underwrite akrasia (for all the reasons we considered in Section II.A).

Now it is usually the case—perhaps it is always the case—that when we find something appealing, we find certain of its features appealing. This invites the thought that finding a thing appealing is just a matter of taking it to have certain appealing properties. But this thought is mistaken. Whether or not taking something to have certain appealing properties is necessary for psychic attraction, it is certainly not sufficient. We can ascribe appealing properties to something without thereby being psychically attracted to it. You have just entered a sauna and begun to

---

95 Here I disagree with Johnston, supra note _, at 190 (asserting that “affect is close to ubiquitous”).
96 Hawkins, supra note _, at 244.
97 Anscombe, supra note _; Raz, Incommensurability, supra note _; Raz, Agency, supra note _.
sweat profusely. You know that within minutes you will have a dry sensation in your throat and that you will be able to alleviate the unpleasant sensation by drinking. Moreover, you are aware that even now, before the dry sensation creeps into your throat, a sip of water would be pleasant and satisfying. Yet you are not currently thirsty: you have no unpleasant sensation in your throat and no felt motivation to drink. Does your mere awareness that drinking would be pleasant constitute a desire to drink? Of course not. At any given moment you may be consciously aware of many possible pleasant experiences and vividly aware of what makes them pleasant, yet not desire any of them. You can be certain that sexual activity would be pleasurable and even represent this to yourself graphically, yet have no desire to engage in it; you can be certain that revenge would be delicious, yet feel no vengeful urge. 98

Another possibility is that psychic attraction consists in selective attention—specifically, selective attention to those very features on account of which we find something appealing. In this vein, consider Scanlon’s notion of a desire in the directed-attention sense: “A person has a desire in the directed-attention sense that P if the thought of P keeps occurring to him or her in a favorable light, that is to say, if the person’s attention is directed insistently toward considerations that present themselves as counting in favor of P.” 99 Scanlon’s notion of a desire in the directed-attention sense gets us closer to what it is to be psychically attracted to something, but not close enough. For we can easily imagine that as you sit in the sauna, your attention is directed insistently toward considerations that render drinking appealing. You cannot but think of how dehydrated you are becoming and of how pleasant it would be to take a sip of water. Yet your throat remains comfortable and you feel no attraction to drinking.

None of this is to deny that when you desire something, your attention is ordinarily directed to appealing properties of the thing you desire. The point is just that desire is what directs your attention. Attention is not what constitutes your desire.

These reflections indicate that our being psychically attracted to something is not just a matter of our ascribing appealing properties to the

98 Cf. Alexander Nehamas, Only a Promise of Happiness: The Place of Beauty in a World of Art 68 (2007) (on “the serious difference between describing a face as attractive and actually being attracted to it”).

99 Scanlon, What We Owe, supra note __, at 39 (emphasis added).
thing, even when we attend to these properties insistently. This seems especially true in cases of romantic or sexual desire. Ascription of appealing properties is obviously insufficient for romantic or sexual desire, since we can recognize that a person has appealing features (and even find these features utterly captivating) without being attracted to the person. What may be less obvious is that the ascription of appealing features is not only insufficient for romantic or sexual desire; it is also unnecessary. When we desire someone romantically, we are often unable to elucidate the basis of our attraction. Though we can usually specify personal attributes that entice us, being attracted to someone is different from being enticed by the person’s attributes. We are attracted to the person as a whole, not to the person qua bearer of certain attributes. Though we can ordinarily answer the question “what do you see in her?” by specifying personal attributes that we genuinely find attractive, these answers tend to miss the point. The same is true for sexual desire. If you are sexually attracted to someone, as opposed to merely judging that she is sexually attractive, this is not a matter of finding certain of her attributes or mannerisms sexually attractive. We are sexually attracted to people, not ordinarily to their features. Being sexually attracted (only) to a person’s features is something close to sexual fetishism, which is quite different from ordinary sexual attraction.

I’ve now considered and rejected four candidates for the role of psychic attraction: (i) felt motivation, (ii) evaluative ‘seeming,’ (iii) ascription of appealing properties to the desired object, and (iv) insistent attention to these properties. Each of these phenomena might ordinarily accompany psychic attraction. Yet none is sufficient for it; thus, none should be thought the essence of desire.

I am not suggesting that psychic attraction could be isolated from all of its usual accompaniments—especially not felt motivation, which seems always to accompany desire. Yet felt motivation is never brute, never the bedrock of desire. On the contrary, we feel motivated to pursue the things we desire because those things psychically attract us. Far from being reducible to motivation, attraction is what grounds felt motivation and what makes it intelligible. From the standpoint of introspection, it is our being psychically attracted to something that makes sense of our being motivated to pursue it, not the other way around.

100 Cf. Nehamas, supra note __, at 99 (‘Even if I love you for your kindness (inadequate as . . . that has to be as a reason), you needn’t fear that I would love someone kinder than you are more than I love you . . . ’).
2. Attraction and Provocation. — What is going through your mind when you are provoked? Ordinarily, there is at least (i) an insistent thought that the provoker has wronged you and deserves to suffer as well as (ii) an intense motivation to strike back. Yet these two psychic states cannot be the whole story, since neither the thought nor the motivation can give rise to intelligible akratic action. So there must be something more.

And there is. Besides the thought and the motivation, there is a desire—the desire to retaliate. I suggest that we understand this desire as an instance of psychic attraction. When you are provoked, you find the prospect of retaliation attractive. You are drawn to it. You find it appealing—even pleasurable. I mean this literally. (It was not an accident that the desires I scrutinized earlier in this Section were almost all desires for various kinds of pleasure.) Poets have long described revenge as “sweet” and neuroscience suggests that this description is the literal truth, or nearly so. When a victim of unfair treatment inflicts so-called altruistic punishment on someone who has wronged him, the victim undergoes brain activity associated with the anticipation of such things as the pleasant taste of sugar. It is not merely that the victim experiences

---

101 The thought that the wrongdoer deserves to suffer need not be a thought that you endorse; in other words, the thought need not constitute a belief. This thought can coexist with a belief to the effect that retaliation is unjustified—a belief that renders akratic your subsequent act of retaliation.

102 See supra Sections II.A and II.B.


105 All the subjects in the study described in de Quervain, supra note _, were male.

106 de Quervain, supra note _,; John P. O’Doherty, et al, Neural Responses during Anticipation of a Primary Taste Reward, 33 Neuron 815 (2002). See also
retaliation as pleasurable; he is *attracted* to the prospect of retaliation in the same way as he would be attracted to the prospect of eating sugar. He is quite literally drawn to retaliation as though it were “sweeter than drops of honey.”  

107 I take this neuroscientific finding to be consistent with the phenomenology of retaliation.  

108 Amid all of the psychic turmoil caused by an outrageous provocation—amid the anger and pain and frustration—there is a powerful current of psychic attraction. We should not assume that this sort of pleasure-directed attraction can occur only in happy times.

The present conception of the retaliatory desire has little trouble explaining how a defendant could satisfy the emotional requirement, which demands that the defendant’s anger be morally understandable in light of the victim’s conduct. Though taking pleasure in another’s suffering is almost certainly not justifiable and probably not even be excusable, it is surely understandable—especially when the other person has wronged you.

The more pressing question is whether the present conception of the retaliatory desire can make sense of volitional impairment. If psychic attraction really is the essence of retaliatory temptation, can retaliatory temptation (so conceived) really give rise to intelligible akratic action? I see no reason why it cannot, as I see no reason to deny that a person can be attracted *most* to something other than what he judges best. Psychic attraction is one thing; judgment is another. Regrettably, these two things do not always go hand in hand.

You may find this answer wholly unsatisfying. You may feel that no matter how we conceive of desire, *akrasia* will remain as mysterious as ever. And indeed it may. But to solve this mystery—if that is really what it is—we need more than a conception of desire. We need a conception of agency. A conception of desire cannot be expected to solve the mystery of *akrasia* by itself. All we can expect of such a conception is that it not render the mystery of *akrasia* insoluble. *Akrasia* may yet be ineluctably mysterious, but that is not because the nature of desire makes it so.

---

107 Homer, supra note _, at XVIII, 109.

Conceiving of desire as psychic attraction does not render *akrasia* mysterious or impossible, the way rival theories do. Recall the difficulties that beset the evaluative conception. First, there is the problematic story that evaluative theorists typically tell about *akrasia*—the story that proposes an analogy between, on the one hand, evaluative appearance and action and, on the other, perceptual illusion and belief. This story ultimately fails to make sense of *akrasia*; I explained, because it fails to make sense of how we could be led astray by evaluative illusions that we recognize as illusions. As I argued, the standard story is worse than unhelpful; it is fatal. Instead of showing how akratic action is intelligible, the standard story seems to entail that akratic action is impossible. If a known perceptual illusion cannot tempt or cause belief, then—taking seriously the analogy between evaluative appearance and perceptual illusion—it would seem to follow that a known illusory evaluative appearance cannot tempt or cause action. Moreover, even if a perceptual illusion could tempt belief, I observed, nothing could be easier than resisting such a doxastic (i.e., belief-related) temptation; and if nothing could be easier than resisting a doxastic temptation, then—taking seriously the analogy between evaluative appearance and perceptual illusion—it would seem to follow that nothing could be easier than resisting a practical temptation.

None of these problems plagues the idea that affective desire consists in psychic attraction, because psychic attraction is fundamentally non-representational. Feeling attracted to an object is just not a matter of representing the object as *being* some way or other. Much less is it a matter of representing the object as being good. How a person represents some object is simply a separate question from whether the object attracts him. So if we conceive of temptation in terms of psychic attraction, we will avoid the seemingly insurmountable challenge of explaining how a person could be led astray by a representation whose content he does not endorse. In other words, we will avoid the problem of explaining how a known illusion could somehow be hard to ignore.

Next, recall the difficulties that beset the mechanistic conception. Like the evaluative conception, the mechanistic conception renders *akrasia* highly problematic. But it does so for a different reason. In this case, the seemingly insurmountable challenge is not the challenge of explaining how resisting temptation might be difficult (the mechanistic conception has no trouble there). Rather, it is the challenge of explaining how a person who is not acting on an irresistible compulsion could in
defiance of his all-things-considered judgment intelligibly pursue what he does not find appealing at all. This challenge arises because the mechanistic conception takes temptation to be nothing but a vector of “sheer brute force,” a raw motivating state that fails to present its object as appealing or attractive. If I were afflicted by such an impulse, and it were not irresistible, what could make sense of my acquiescing? Consistent with the mechanistic conception, we cannot appeal to the pleasure of acquiescence, because this pleasure would make sense of my acquiescing only if we understood my very attraction to the pleasure non-mechanistically. Nor can we appeal to the displeasure of resistance, because this displeasure would make sense of my acquiescence only if we understood my aversion to the displeasure non-mechanistically. On pain of rendering akrasia unintelligible, the mechanistic conception must therefore suppose temptation to be irresistible. Thus, contrary to what the provocation defense presupposes, the mechanistic conception will ultimately entail that the provoked defendant’s loss of self-control is fully excusable.

If we conceive of temptation in terms of psychic attraction, then there is no barrier to our accepting that resisting temptation is not ordinarily impossible. Temptation qua psychic attraction need not be irresistible in order for our acquiescence to be intelligible. Indeed, it is often the case that we succumb to temptation not because resistance would be futile but because it would be no fun.

CONCLUSION

I began by considering the two dominant approaches to understanding why a killer’s passion sometimes lessens the gravity of her crime—the volitional theory, which assesses the passion’s relationship to the killer’s subsequent loss of self-control, and the evaluative theory, which assesses the passion’s relationship to the victim’s antecedent act of provocation. I argued that each of these two theories contains a seed of truth. Properly understood, the doctrine of provocation submits the killer’s psychological condition to two levels of normative scrutiny: looking backward, the doctrine asks whether the defendant’s anger was morally understandable in light of the victim’s prior conduct; looking forward, the doctrine asks whether the defendant’s anger was apt to cause a morally understandable loss of self-control. Then I offered a theory of affective desire that shows how a person’s anger could be both of these things at once.