

# COLUMBIA LAW REVIEW

VOL. 96

MARCH 1996

NO. 2

## TWO CONCEPTIONS OF EMOTION IN CRIMINAL LAW

*Dan M. Kahan\**  
and *Martha C. Nussbaum\*\**

Introduction .....	270
I. Two Conceptions of Emotion .....	275
A. The Mechanistic Conception .....	278
1. The Basic Account .....	278
2. History .....	280
3. Arguments Against the Mechanistic Conception ....	282
B. The Evaluative Conception .....	285
1. The Basic Account .....	285
2. History and Development .....	289
3. Different Species of the Evaluative Conception ....	293
4. The Role of Social Factors .....	296
C. Emotions, Character, and Moral Education .....	297
II. Two Conceptions of Emotion in Substantive Criminal Law ..	301
A. Emotions and Accountability .....	301
B. Voluntary Manslaughter .....	305
1. The Common Law Formulation .....	305
a. Adequate Provocation .....	306
b. Mitigation vs. Exculpation .....	312
c. Heat of Passion .....	315
d. Cooling Time .....	316
e. Justification or Excuse? .....	318
2. The Model Penal Code Formulation .....	321
C. Premeditated Murder .....	323
D. Self-Defense .....	327
E. Duress .....	333
F. Voluntary Act .....	338
G. Insanity .....	341

\* Assistant Professor, University of Chicago Law School. Professor Kahan is grateful to the Russell J. Parsons Fund for Faculty Research and to the Morton C. Seeley Endowment Fund at the University of Chicago Law School for financial support.

\*\* Professor of Law and Ethics, University of Chicago. The authors are grateful to Albert Alschuler, Richard Craswell, Richard Epstein, David Friedman, Stephen Gilles, Peter Huang, Lawrence Lessig, Tracey Meares, Daniel Meltzer, Richard Posner, Stephen Schulhofer, David Strauss, and Cass Sunstein for written comments on an earlier draft; to participants in workshops sponsored by the Brown University Philosophy Department, Harvard Law School, the University of Chicago Law School, the University of Michigan Philosophy Department, and the R.I.S.K. Project at the University of Iowa for oral comments; and to Sally Abrahamson, Ross Davies, Timothy Delaune, and Jennifer Wisner for research assistance.

III. Shifting Assessments of Emotion in Criminal Law.....	346
IV. Defending the Evaluative Conception of Emotion in Criminal Law .....	350
A. The Evaluative Conception and the Purposes of Criminal Law.....	350
1. Expressive Condemnation. ....	351
2. Optimal Deterrence.....	353
a. Optimal Regulation of Emotional Behavior.....	354
b. Formation of Appropriate Emotional Dispositions .....	355
3. Individual Desert .....	357
B. Moral Dilemmas.....	359
1. Is the Evaluative Conception Illiberal? .....	359
2. The Problem of "Bad Morality" and a (Partial) Institutional Solution.....	362
a. No Worse, Possibly Better.....	362
b. A (Partial) Institutional Solution .....	364
C. Conviction and Sentencing: A Role for Mercy? .....	366
Conclusion .....	372

#### INTRODUCTION

Emotions are ubiquitous in criminal law, as they are in life. But how do they, and how should they, affect legal assessment? Should the law be more sympathetic to defendants who are taken over by passions such as anger and fear, or should it view such defendants as especially dangerous? Or should the response of the law depend on an appraisal of the emotion itself—whether it is appropriate or inappropriate, "reasonable" or "unreasonable"? What does it mean for an emotion to be reasonable? Aren't emotions, after all, just disturbances of the personality that can be more or less strong but that are always hostile to reason? Or do they embody judgments, ways of seeing the world? If they do, should we hold people morally accountable for those judgments?

The law does not appear to give consistent answers to these questions. Consider:

¶ Frank Small had a quarrel with C.R. Jacoby in Keyser's Saloon. Jacoby walked out of the saloon and down the street with his wife. As he was walking away, Small came up to him, put a pistol to his head, and shot him. Jacoby died two days later. In an attempt to mitigate the grade of the homicide to manslaughter, Small argued that he had been impelled to kill by an intense flash of anger that failed to abate in the minutes between the quarrel and the fatal attack; on appeal from his conviction for first degree murder, he argued that the trial court had erred in failing to instruct the jury that the "space of time which will be deemed sufficient for a man to cool, after a conflict, may differ with different persons." The Pennsylvania Supreme Court rejected this contention: "Suppose then we admit testimony that the defendant is quick-tempered, violent and re-

vengeful; what then? Are these an excuse for, or do they even mitigate crime? Certainly not, for they result from a want of self-discipline; a neglect of self-culture that is inexcusable."<sup>1</sup>

¶ Robert Elliott broke into his brother's home and shot him dead for no apparent reason. At his trial, the defense introduced psychiatric testimony that Elliott was acting "under the influence of an extreme emotional disturbance caused by a combination of child custody problems, the inability to maintain a recently purchased home and an overwhelming fear of his brother." Instructed to convict Elliott of manslaughter only if the brother had adequately provoked him, the jury found him guilty of murder. But the Supreme Court of Connecticut reversed: whether or not Elliott was provoked, the jury was free to convict of the lesser charge once it determined that his "self-control[] and reason [were] overborne by intense feelings such as passion, anger, distress, grief, excessive agitation or other similar emotions."<sup>2</sup>

¶ "Babe" Beard was returning home when he encountered a group of young men in a field making away with his cow. Following an angry exchange of words, one of the men approached Beard, saying, "Damn you. I will show you." An enraged Beard then smashed the man's head with the butt of his shotgun, inflicting a fatal wound. The jury convicted Beard after being instructed that it could credit his self-defense claim only if it found that he had no opportunity to withdraw from the confrontation. The Supreme Court of the United States reversed, concluding that a "true man" has no duty to retreat when confronted by wrongful aggression. Beard "was entitled to stand his ground and meet any attack made upon him," rather than endure the shame of flight.<sup>3</sup>

¶ Judy Norman had for years been physically and mentally abused by her husband, who forced her to engage in prostitution and who frequently threatened to kill her. One evening, after a particularly severe beating and after her husband had called her a "dog" and made her lie on the floor while he lay on the bed, Norman took the baby to her mother's house and returned with a pistol. She shot her husband, fatally, while he slept. At trial, a defense expert testified that Norman killed because she feared that if she did not she would "be doomed . . . to a life of the worst kind of torture and abuse" and that "escape was totally impossible." The North Carolina Supreme Court affirmed the trial court's refusal to instruct on self-defense. The majority opinion held that the evidence "would not support a finding that the defendant killed her husband due to a reasonable fear of imminent death or great bodily harm"; the dissent countered that the husband's "barbaric conduct . . . reduced the quality of the defendant's life to such an abysmal state that

---

1. *Small v. Commonwealth*, 91 Pa. 304, 306, 308 (1879).

2. *State v. Elliott*, 411 A.2d 3, 5, 8 (Conn. 1979).

3. *Beard v. United States*, 158 U.S. 550, 552, 561, 564 (1895).

... the jury might well have found that she was justified in acting ... for the preservation of her tragic life."<sup>4</sup>

¶ Claudia Brenner and Rebecca Wight were hiking along the Appalachian Trail in Adams County, Pennsylvania. They found an appropriate campsite for the night, and were making love when Stephen Carr, a stranger who had been lurking in the woods, shot them, killing Wight and striking Brenner five times in the head, neck, and arm. Carr argued at trial that the sight of their lovemaking provoked an uncontrollable sense of revulsion and disgust; in support of this defense—which he argued mitigated the homicide to manslaughter—Carr proffered psychiatric evidence relating to his rejection by women, including his mother, whom he suspected of being a lesbian. The trial court refused to admit this evidence, and Carr was found guilty of first-degree murder. The Pennsylvania Superior Court affirmed, holding that "[the law] does not recognize homosexual activity between two persons as legal provocation sufficient to reduce an unlawful killing ... from murder to voluntary manslaughter." "A reasonable person," the Court concluded, "would simply have discontinued his observation and left the scene; he would not kill the lovers."<sup>5</sup>

¶ David Thacker met Douglas Koehler in a bar and invited him to his apartment. There, Koehler allegedly attempted to kiss Thacker, who became enraged and insisted that Koehler leave. Later in the evening, still angry, Thacker recruited his roommate to help him track Koehler down. When they found him, Thacker shot Koehler in the face, killing him. Outraging gays, prosecutors permitted Thacker to plead guilty to manslaughter, and the judge sentenced him to a mere six years. Explaining the lenient sentence, the judge stated that the unusual circumstances of the killing made it "a one-time tragedy" and that he was "confident Mr. Thacker would not kill again."<sup>6</sup>

As even this small group of cases indicates, the law has treated emotions in several different and conflicting ways. Indeed, it is virtually inconceivable that any consistent theory of what emotions are and why they matter could have generated these results. If emotions are in fact nonreasoning forces that impair self-control, aren't Small and Carr as entitled to mitigation as Elliott? On the other hand, if they are elements of the personality that are subject to "self-culture," and that can be either reasonable or unreasonable, shouldn't Elliott and Thacker be as severely condemned as Small and Carr? What, on this account, makes Beard's

---

4. *State v. Norman*, 378 S.E.2d 8, 9, 11, 13 (N.C. 1989); *id.* at 17, 21 (Martin, J., dissenting).

5. *Commonwealth v. Carr*, 580 A.2d 1362, 1363–65 (Pa. Super. Ct. 1990). See generally Claudia Brenner, *Eight Bullets: One Woman's Story of Surviving Anti-Gay Violence* (1995) (Brenner's personal account of the shooting and its aftermath).

6. *Victim's Family, Gays Say Killer Got Off Too Easy*, Salt Lake Tribune, Aug. 16, 1994, at C1; *Judge Draws Protest After Cutting Sentence of Gay Man's Killer*, N.Y. Times, Aug. 17, 1994, at A15.



fear of being perceived a coward reasonable, and Norman's fear of personality-destroying violence and degradation unreasonable? Perhaps, in contrast to both of these approaches, the law should ask only whether the circumstances in which a person experienced an emotion reveal him to be sufficiently dangerous to warrant incapacitation. But then in what sense was Thacker any less dangerous than Carr? Can the law possibly identify which impassioned offenders should be regarded as dangerous without taking contentious positions on whether their emotions are appropriate or inappropriate to their situation?

We will argue that the disparate approaches to emotion at work in the criminal law stem from a long-standing dispute in Western culture about the nature and educability of the emotions. In this history, two views compete to explain such experiences: what we shall call the mechanistic and the evaluative conceptions of emotion. The mechanistic conception sees emotions as forces that do not contain or respond to thought; it is correspondingly skeptical about both the coherence of morally assessing emotions and the possibility of shaping and reshaping persons' emotional lives. The evaluative conception, in contrast, holds that emotions express cognitive appraisals, that these appraisals can themselves be morally evaluated, and that persons (individually and collectively) can and should shape their emotions through moral education. For decades, philosophers, anthropologists, and psychologists have struggled to get to the bottom of the disagreement between the mechanistic and evaluative accounts. As the conflicting results in these cases suggest, it is important that those who administer the criminal law get to the bottom of it as well.

Our primary claims are descriptive. We believe that it is much easier to see where and how the law is being inconsistent once the evaluative and mechanistic conceptions of emotion are brought into view. Discrepancies in the definition of voluntary manslaughter, for example, stem from an outright conflict between the two conceptions. Other inconsistencies are internal to one conception or the other. The mechanistic conception of emotion often supports different results depending on whether it is combined with a voluntarist or a consequentialist account of moral responsibility. The evaluative view, too, can generate different results—on, for example, the scope of self-defense—depending on what emotional evaluations the law is prepared to endorse.

Recognizing the two conceptions also helps to explain historical shifts in the law's treatment of emotions. The mechanistic conception's influence on the law is a relatively recent phenomenon. Nevertheless, it would be a mistake to conclude that the mechanistic view has eclipsed the evaluative. On the contrary, the evaluative view remains dominant in legal reasoning (if not in academic commentary) and has in fact repelled advances by the mechanistic conception on several doctrinal fronts.

What has changed dramatically, however, is the content of the law's own evaluations. The law's approach to emotional behavior is widely per-

ceived to be in a state of flux.<sup>7</sup> Some see in this an emerging spirit of lawlessness.<sup>8</sup> But with the aid of the evaluative view, we see something else: the historical responsiveness of the law to changes in, and dissensus over, social norms. As the norms that define what kinds of goods are valued by a reasonable person change, so too does the law's appraisal of emotions that reflect such valuations. The law's emerging (if contested) receptiveness to the fear of the domestic violence victim—and its receding sympathy to the rage of the cuckold—stem from changes in norms relating to gender and equality. The power of the evaluative view to explain this phenomenon is another major descriptive benefit of this account of emotion in criminal law.

In addition to describing the law, we also wish to assess it. Our normative claim, simply put, is that doctrines informed by the evaluative conception are superior to ones informed by the mechanistic conception. They are better, in part, because they are more effective in deterring and condemning criminal wrongdoing. But even more important, they are better because they are brutally and uncompromisingly honest. Mechanistic doctrines, by contrast, tend to disguise contentious moral issues. This is bad. The law is more likely to be just, we argue, when decisionmakers are forced to take responsibility for their appraisals of wrongdoers' emotions, and when the public is allowed to see for itself the appraisals that its decisionmakers have made.

Our analysis has four parts. In Part I, we present an overview of the mechanistic and evaluative conceptions of emotion. In addition to identifying the salient characteristics of the two views, we trace their historical origins, their influence on various schools of psychology and philosophy, and their connection to different theories of character and moral education.

In Parts II and III, we develop our descriptive claims. Part II uses the mechanistic and evaluative conceptions of emotion to explain the contours of a series of substantive criminal law doctrines, including voluntary manslaughter, premeditated murder, self-defense, duress, voluntary act, and insanity. We show that numerous disagreements about the structure and purpose of these doctrines track the tension between the mechanistic and evaluative views. Indeed, we argue that a major deficiency in existing academic accounts—whether rooted in individual desert, deterrence, or other perspectives—is that they have not properly taken account of the influence of the evaluative view in shaping criminal law doctrine. Part III extends our descriptive analysis, using the evaluative view to explain historical shifts in the law's assessment of emotions.

---

7. See, e.g., Stephen J. Morse, *The "New Syndrome Excuse Syndrome,"* *Criminal Just. Ethics*, Winter/Spring 1995, at 3-4; Stephen J. Schulhofer, *The Trouble with Trials; the Trouble with Us*, 105 *Yale L.J.* 825, 850-51 (1995).

8. See, e.g., Alan M. Dershowitz, *The Abuse Excuse* 3-5 (1994); Charles J. Sykes, *A Nation of Victims: The Decay of the American Character* (1992).

Finally, Part IV normatively appraises the mechanistic and evaluative conceptions of emotions as they appear in criminal law. We argue that doctrines structured to reflect the evaluative view best promote the recognized purposes of criminal law. We also address, and rebut, the charge that evaluation of emotions commits the law to making inappropriate moral judgments. Finally, we qualify our support of the evaluative conception of emotion in criminal law by defending the practice of *mercy* in the sentencing of offenders who cannot justly be held responsible for their characters.

### I. TWO CONCEPTIONS OF EMOTION

Since the time of the ancient Greek philosophers, the mechanistic and evaluative conceptions of emotion have competed in the Western philosophical tradition, and in popular views deriving from and also influencing that tradition.<sup>9</sup> More recently, the debate has been taken up in psychology and anthropology as well.<sup>10</sup>

Before we describe these views, we must describe more clearly what they attempt to explain. Throughout this tradition, writers have generally agreed that certain human experiences that people commonly call "emotions"<sup>11</sup> can usefully be classified together, since they share many com-

---

9. Compare Martha C. Nussbaum, Poetry and the Passions: Two Stoic Views, in *Passions & Perceptions: Studies in Hellenistic Philosophy of Mind* 97, 104–22 (Jacques Brunschwig & Martha C. Nussbaum eds., 1993) [hereinafter Nussbaum, Poetry and the passions] (contrasting "cognitive" and "non-cognitive" Stoic views of the passions) with Ronald de Sousa, The Rationality of Emotions, in *Explaining Emotions* 127, 131 (Amélie O. Rorty ed., 1980) (discussing the divergent views among both ancient and modern philosophers regarding the rationality or irrationality of emotions, and attempting to answer the question, "Which sort of rationality can emotions aspire to?"). Insofar as we have been able to investigate non-Western traditions, they seem to conform strongly to the evaluative conception. For good studies of Indian and Chinese views, see *Emotions In Asian Thought* (Roger T. Ames & Joel Marks eds., 1994). Kwasi Wiredu, the leading American expert in the history of African philosophy, told one of the authors in conversation that in the Ghanaian tradition, which he knows best, there would be no way at all to speak of "emotions" as a class distinct from thoughts or judgments; they would simply be described as thoughts of a certain sort. Conversation with Kwasi Wiredu, Professor of Philosophy, University of South Florida, at Duke University, in Durham, N.C. (Mar. 24, 1994).

10. See, e.g., Richard S. Lazarus, *Emotion & Adaptation* 8–15 (1991) (describing "The Retreat from Radical Behaviorism and the Rise of Cognitivism" in the 1960s and 1970s).

11. In much of the debate, "passions" (compare Greek *pathê*, Latin *passiones*, French *passions*) appears, sometimes alongside "emotions." (Latin *passio* is not in common use in the Classical period: Cicero and Seneca translate Greek *pathê* by *affectus*, which correctly renders the link with the verb *paschô*, which in general means "to be affected." *Affectus* is also Augustine's chosen term, along with *perturbationes* and *passiones*. See Saint Augustine, *The City of God Against the Pagans*, Books XIV.9, .10, .23, at 310, 322, 382 (Philip Levine trans., 1966) (citations to *City of God* include the standard book and chapter designations followed by the page numbers in the Levine edition).) Descartes uses French *passions* as the generic term, see Descartes, *Les passions de l'âme*, reprinted in *Œuvres et Lettres* 695 (André Bridoux ed., 1953) (1649), but uses the term *émotions* for some species. Id. at 732,

mon features. The major emotions include joy, grief, fear, anger, hatred, pity or compassion, envy, jealousy, hope, guilt,<sup>12</sup> gratitude, disgust, and love. Philosophers, psychologists, and anthropologists generally agree that these are distinct, in important respects, from bodily appetites such as hunger and thirst, and also from objectless moods, such as irritation or endogenous depression. There are many distinctions among members of the family;<sup>13</sup> the classification of some cases remains a matter of dispute;<sup>14</sup> but there is still great consensus about the central members in the family and their distinctness from other human experiences.

---

738. The term "passion" by now suggests "a vehement, commanding, or overpowering emotion." The Oxford English Dictionary 309 (J.A. Simpson & E.S.C. Weiner eds., 2d ed. 1989), but that was not the case with its earlier generic uses in Greek, Roman, and French philosophy. In general, even when the generic term used fluctuates, the major species remain remarkably consistent.

12. Guilt was in fact not included in any ancient list of the emotions; ancient taxonomies recognized only present-directed and future-directed emotions. See 3 Hans von Arnim, *Stoicorum Veterum Fragmenta* 397, 401, 409, 414, at 96-100 (1924) (collecting the ancient lists) (citations to the Stoic fragments include the fragment number followed by the page numbers in the von Arnim text).

13. For example, love is frequently analyzed as a relationship involving emotion, not simply as an emotion. See, e.g., Aristotelis, *Ethica Nicomachea* 1156a-1156b, at 158-60 (Ingram Bywater ed., 1949) (1894) [hereinafter Aristotle, *Ethics*] (citations to the *Ethics* include the standard "Bekker" section designations followed by the page numbers in the Bywater edition) ("for the greater part of the friendship of love depends on emotion"); Martha C. Nussbaum, *Eros and the Wise: The Stoic Response to a Cultural Dilemma*, in 13 *Oxford Studies in Ancient Philosophy* 231, 233 (1995) [hereinafter Nussbaum, *Eros and the Wise*] ("*erós* is seen as a divine gift, connected with . . . generous educative intentions toward the beloved"). Hatred may have the same features, although it has also been analyzed simply as an emotion. Compare W. Ronald D. Fairbairn, *Psychoanalytic Studies of the Personality* 26-27 (1984) (treating hate as the "reversal of values" of love in individuals with a schizoid tendency) and Nico H. Frijda, *The Emotions* 212 (1986) ("*Hatred* is an emotion that contains the component of object evaluation.") with Descartes, *The Passions of the Soul*, in 1 *The Philosophical Writings of Descartes* 325, 350 (John Cottingham et al. trans., 1985) (classifying hatred as a "passion[ ]") and Sigmund Freud, *A General Introduction to Psycho-Analysis* 293 (Joan Riviere trans., rev. ed. 1935) [hereinafter Freud, *Psycho-Analysis*] (characterizing hatred as "actuate[d by] a feeling of aversion"). Some varieties of disgust seem more like instinctual bodily reactions than like emotions, though many varieties lie closer to emotions in containing a view of an object. See, e.g., Lazarus, *supra* note 10, at 56 (contrasting "distaste" as a built-in reflex with "disgust" as a learned response to a particular "substance, idea, or action").

14. Usually this uncertainty about classification derives from an uncertainty about the salient features of the particular appetite or emotion or mood in question—about sexual desire, for example, as to whether it is primarily a bodily urge, or whether it is aroused by focusing on an object; about depression, as to whether it always has some object or reason, or whether it can arise from purely endogenous physiological causes. See Martin E.P. Seligman, *Helplessness: On Depression, Development, and Death* 78-79 (1975) (summarizing the "endogenous-reactive dichotomy" in analysis of depression); George Graham, *Melancholic Epistemology*, 82 *Synthese* 399, 403-07 (1990) (describing contrasting cases). Other borderline cases that can be found in some classifications are pride (is this really an emotion, or a trait of character?), humility (*ditto*), admiration (emotion, or nonemotional thought?), and respect (*ditto*).

Philosophical accounts of emotion have their roots in ordinary ways of talking and thinking about the emotions. It is always important to thinkers on both sides of the debate to point to the (alleged) intuitive experiential credentials of their view, and therefore to cite common sayings, poetic descriptions, and so forth, as evidence supporting their side.<sup>15</sup> Although these real life experiences certainly exhibit some historical variation (love, for example, appears in subtly different forms across the ages, from ancient Greek *erôs* to courtly love to nineteenth-century romantic conceptions, and so forth),<sup>16</sup> there is enough common ground and overlap that we are entitled to think of the debate as a genuine debate about something reasonably stable, rather than as a set of descriptions of a changing reality.<sup>17</sup>

Because the philosophical accounts track common perceptions of emotions in this way, our claims about the law need not rest on claims of direct theoretical influence. Nevertheless, to understand the situation in the law, it is useful to study the more explicit theoretical debates about emotion, traces of which appear in legal writings, often in compressed and unacknowledged form.

To introduce the two sides in a highly schematic way, we may say that the mechanistic view holds that emotions are forces more or less devoid

15. Compare Seneca, *On Anger*, in *Seneca: Moral and Political Essays* 1.12.1–2, at 1, 30 (John M. Cooper & J.F. Procopé eds. & trans., 1995) [hereinafter *Seneca, On Anger*] (citations to *On Anger* include the standard section designations followed by the page numbers in the Cooper edition) (“‘My father is about to be killed—I will defend him; he has been killed—I will avenge him; not because I am pained, but because I should.’”) and id. at 1.16.7, at 34–35 (quoting Zeno analogizing emotions to wounds) with Aristotle, *Ethics*, supra note 13, at 1155a32–1155b1, at 156 (illustrating elements of love and friendship with such phrases as “birds of a feather flock together” and “two of a trade never agree”). Even the most elaborately theorized accounts, for example, ancient Greek Stoic taxonomies, are carefully buttressed by appeals to common usage and experience. See Martha C. Nussbaum, *The Therapy of Desire* 368–69 (1994) [hereinafter *Nussbaum, Therapy*].

16. See, e.g., Beth L. Bailey, *From Front Porch to Back Seat: Courtship in Twentieth-Century America* 1–12 (1988) (chronicling the twentieth-century transformation of courtship into “a private act conducted in the public world”); Henry Staten, *Eros in Mourning: Homer to Lacan* 98–107 (1995) (analyzing treatment of courtly love by Shakespeare); John J. Winkler, *The Constraints of Desire: The Anthropology of Sex and Gender in Ancient Greece* 72, 82 (1990) (contrasting ancient Greek eros and modern conceptions of love); John F. Benton, *Clio and Venus: An Historical View of Medieval Love*, in *The Meaning of Courtly Love: Papers of the First Annual Conference of the Center for Medieval and Early Renaissance Studies* 19, 19–37 (F.X. Newman ed., 1968) (reviewing various treatments of love in medieval times, and criticizing modern use of the term “courtly love”); David M. Halperin, *Plato and the Metaphysics of Desire*, in *5 Proceedings of the Boston Area Colloquium in Ancient Philosophy* 27, 27–36, 50–51 (John J. Cleary & Daniel C. Shartin eds., 1989) (describing and analyzing Plato’s theory of erotic desire); Nussbaum, *Eros and the Wise*, supra note 13, at 231–40 (summarizing ancient Greek eros).

17. We do not mean here to deny that emotions exhibit considerable social variation and are in that sense to a degree “socially constructed.” See *infra* text accompanying notes 107–111.

of thought or perception—that they are impulses or surges that lead the person to action without embodying beliefs, or any way of seeing the world that can be assessed as correct or incorrect, appropriate or inappropriate. The evaluative view holds, by contrast, that emotions do embody beliefs and ways of seeing, which include appraisals or evaluations of the importance or significance of objects and events. These appraisals can, in turn, be evaluated for their appropriateness or inappropriateness.<sup>18</sup>

This Part describes the two views and some of the arguments by which each defends itself. We shall spend more time on criticisms of the mechanistic conception made by the evaluative conception, since we want to show why the evaluative conception has enjoyed widespread acceptance and has proven the dominant influence in the development of law, if not in legal scholarship. This influence cannot be altogether separated from the factors that make it a powerful account of some aspects of human experience, and an effective basis for practices of moral education and moral appraisal. Thus, in the course of describing the evaluative conception, we will also be arguing for it, although we do not intend the account that follows to be a full defense of the evaluative view.

#### A. *The Mechanistic Conception*

1. *The Basic Account.* — The mechanistic conception has appeared in many forms. Words associated with it include “impulse,”<sup>19</sup> “drive,”<sup>20</sup> “unreasoning movement,”<sup>21</sup> “force,”<sup>22</sup> and no doubt others. The basic claim of this conception is that emotions (for example, anger or fear) are energies that impel the person to action, without embodying ways of thinking about or perceiving objects or situations in the world. They are not even very reliably hooked up to these ways of thinking and seeing: they follow laws of their own.<sup>23</sup> The mechanistic conception typically sees these laws as deriving from an innate “human nature” that precedes so-

---

18. This account is closely connected to the historical argument in Nussbaum, *Therapy*, *supra* note 15, chs. 10–11; and to the more elaborate philosophical argument in Martha C. Nussbaum, *Upheavals of Thought: A Theory of the Emotions*, The Gifford Lectures for 1993 (forthcoming 1998) [hereinafter Nussbaum, *Upheavals*].

19. See, e.g., Richard Wollheim, *Sigmund Freud* 111 (1971) (discussing “the taming of impulse”). For contrasting views of “impulse” in the courts, compare *McGill v. State Accident Ins. Fund Corp.*, 724 P.2d 905, 906–07 (Or. Ct. App. 1986) (treating a suicide as an irresistible impulse) with *Kwosek v. State*, 100 N.W.2d 339, 346 (Wis. 1960) (“Such so-called irresistible impulses are merely unresisted urges.”).

20. See, e.g., Immanuel Kant, *Introduction to The Metaphysics of Morals*, in *Kant’s Ethical Philosophy* 1, 14 (James W. Ellington trans., 1983); see also Lazarus, *supra* note 10, at 8 (listing and criticizing mechanistic views).

21. See Galen, *On the Doctrines of Hippocrates and Plato* 332 (Phillip de Lacy ed. & trans., 1984) (summarizing Posidonius).

22. See, e.g., Freud, *Psycho-Analysis*, *supra* note 13, at 306.

23. See Nussbaum, *Poetry and the passions*, *supra* note 9, at 105.

cial shaping and responds to social cultivation only to a limited degree.<sup>24</sup> Emotions lie behind culture; they are parts of our basic innate human equipment, to be studied by the sciences of psychology and physiology<sup>25</sup> rather than by the normative disciplines of ethics and political thought. Characteristically, this view refuses to make a strong distinction between emotions such as fear, grief, anger, and envy, and bodily appetites such as hunger and thirst.<sup>26</sup> Like these appetites, emotions are seen as feelings relatively devoid of representational or cognitive content. Like these appetites, emotions may have an object, but the object is conceived of as an external cause or trigger of the emotion, rather than as something focused on within the emotion itself.<sup>27</sup>

The mechanistic conception has force because it appears to capture well some prominent features of emotional experience. First, it captures a connection between emotion and passivity that occurs in much of our talk and experience. Emotions feel like things that sweep over us, or sweep us away, or invade us, often without our consent or control—and this intuitive idea is well preserved in the view that they really are impulses or drives that go their own way without embodying reasons or be-

24. See, e.g., Immanuel Kant, *Perpetual Peace*, in *Kant: Political Writings* 93, 111, 120 n.\* (Hans Reiss ed. & H.B. Nisbet trans., 2d ed. 1991) (describing war as “ingrained in human nature,” and concluding that international strife is “irrefutabl[e]” evidence of inherent human wickedness); see also Catherine A. Lutz, *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll & Their Challenge to Western Theory* 53–80 (1988) (reviewing and critiquing Western views of emotions).

25. See, e.g., Max F. Meyer, *That Whale Among the Fishes—The Theory of Emotions*, 40 *Psychol. Rev.* 292, 292–97 (1933).

26. See, e.g., Plato, *The Republic* 437–42 (Johannes Burnette ed., 1954) (1902) (citations to *The Republic* indicate the standard “Stephanus” section designations); see also Galen, *supra* note 21, at 322.

27. See, e.g., Sigmund Freud, *Three Essays on the Theory of Sexuality* 32–34 (James Strachey ed. & trans., Basic Books rev. ed. 1975) (describing, in instinctual terms, the transformation of love into hate, and characterizing “instinct” as “the psychical representative of an endosomatic, continuously flowing source of stimulation”); Melanie Klein, *On Identification* (1955), reprinted in 3 *Writings of Melanie Klein: Envy and Gratitude and Other Works 1946–1963*, at 141, 152–54 (Roger Money-Kyrle et al. eds., 1975) (describing the “emotions” of a fictional patient as responses to, for example, parental “primal objects” and internalized “good” and “bad objects”). The contest between the two views has nothing to do with the contest between dualist and physicalist philosophies of the mind-body relation. One may accept the mechanistic view and yet not accept a physiological account of the impulses or forces in question. See, e.g., Freud, *Psycho-Analysis*, *supra* note 13, at 180 (attributing “impulses of hate” to “recollections”). One may also hold that all processes in living creatures are bodily without concluding that this licenses the elimination of perception and thought from our definitions of emotion. See, e.g., Julia Annas, *Hellenistic Philosophy of Mind* ch. 2 (1992) (analyzing the Stoics’ inclusion of perception and thought in their conception of a “physical” soul); Martha C. Nussbaum & Hilary Putnam, *Changing Aristotle’s Mind*, in *Essays on Aristotle’s De Anima* 27, 44 (Martha C. Nussbaum & Amélie O. Rorty eds., 1992) (asserting that “[t]he proper conclusion, Aristotle’s conclusion, is that all these, perceiving, desiring, emotion, are formulae in matter” and attributing “intentional awareness” and “perception” to at least some emotion).

liefs.<sup>28</sup> Second, the view captures a sense we have that emotions are external to the self, forces that do something to "us" without being (or at least without clearly being) parts of what we think of as ourselves. Anger, for example, can seem to come boiling up from nowhere, in ways of which "we" strongly disapprove. Finally, the view appears to capture the urgency and "heat" of the emotions, the sense we have that they do have enormous force—for if we think of them as drives or forces similar to currents of an ocean, we can imagine these natural forces as extremely strong without being troubled by questions about how our own thoughts could have such force. To many thinkers, seeing emotions this way has appeared to be both intuitively plausible and appropriately scientific; we study them the way we would study a jolt of electricity going through an organism.<sup>29</sup>

Any view that opposes this conception will have to find some way of explaining these intuitive phenomena. Conceptions that define emotions as embodying a kind of thought about an object would appear to have difficulty meeting this challenge, for thoughts are usually seen as things we actively make or do, not things we suffer; they are usually conceived of as central to the core of our selfhood; and they are usually imagined as calm and cool. Thinking of emotions as thoughts may make it difficult to see why they should be so difficult to manage and should cause the upheaval in human life that they frequently do.

2. *History.* — The mechanistic conception has had a long history in Western philosophy, especially in thinkers strongly influenced by the rise of mechanistic and scientific conceptions of the human being. Although the view can be found in classical antiquity—in some writings of Plato,<sup>30</sup> for example, and in the work of the medical philosopher Galen<sup>31</sup>—it becomes popular with the advent of mechanistic explanations of human life generally, in the work of Descartes<sup>32</sup> and, in a different way, the British Empiricists.<sup>33</sup> Kant, probably under the influence both of Christian no-

---

28. An important variant of this intuitive idea—and one whose influence on the history of law and popular conceptions should not be underestimated—is the Christian idea of original sin, as discussed, for example, in Saint Augustine's works, in which emotions are seen as ungovernable relics of an innate nature that take us over in our sleep even if we manage to keep them down during our waking life. See, e.g., 2 Saint Augustine, *St. Augustine's Confessions* 150–52 (T. E. Page et al. eds. & William Watts trans., 1961).

29. See, e.g., Lutz, *supra* note 24, at 65–66 (reviewing psychological and philosophical treatments of emotion as a physical force).

30. See, e.g., Plato, *supra* note 26, at 439–41.

31. See, e.g., Galen, *supra* note 21, at 249.

32. See generally Descartes, *supra* note 13, at 328 (developing a theory of the passions "as if I were considering a topic that no one had dealt with before me" because of "the defects of the sciences we have [received] from the ancients"). On the complexities of Descartes's position, see the excellent account in Anthony Kenny, *Action, Emotion, and Will* 1–17 (1963).

33. On David Hume, see Kenny, *supra* note 32, at 20–28. Donald Davidson has stressed the complex affiliations of Hume's view with the evaluative, as well as the mechanist, traditions:



tions of our original nature and of German Romantic ideas of a pre-cultural nature, endorses a form of the mechanistic conception.<sup>34</sup> His major role in the history of political and legal thought makes his endorsement of the mechanistic view particularly important to our argument.

In this century, the mechanistic conception has enjoyed great prominence, owing to the influence both of these philosophical views and of several different forms of psychological theory. Until rather recently, cognitive psychologists under the influence of behaviorism held that emotions, like other psychological states, could be understood as impulses or forces, in this case bodily forces, that do not contain thoughts or evaluations.<sup>35</sup> Although in most cases they did not offer concrete ideas of how to reduce emotions to noncognitive physiological states, they persisted in their confidence that this could ultimately be done.<sup>36</sup> Meanwhile, in psychoanalysis, Freud and those who most closely followed him offered a mechanistic account of both appetites and emotions, treating emotions as powerful impulses or drives that are innate and not very responsive to information about the world.<sup>37</sup> Although the mechanistic conception soon came under sharp theoretical attack in both cognitive psychology and psychoanalysis (for example by the "object-relations"

---

Hume will not allow that pride or any other passion is based on reason alone; but this is not to deny that some passions . . . are based on *reasons*. Hume's point is rather that if a passion is based on reasons . . . at least one of the reasons must itself be, or be based on, a passion.

Donald Davidson, *Hume's Cognitive Theory of Pride*, 73 J. Phil. 744, 752 (1976). But see Annette Baier, *Hume's Analysis of Pride*, 75 J. Phil. 27, 27-40 (1978) (criticizing Davidson's account).

34. Kant's moral theory does not entail the mechanistic view, and in some respects would be better supported by the evaluative view. See Nancy Sherman, *Making a Necessity of Virtue: Aristotle and Kant on Virtue* (forthcoming 1996) (manuscript at 174, on file with the Columbia Law Review) (noting Kant's distinction between the source of morality's constraints—practiced reason—and the "actual execution" of moral principles); Martha Nussbaum, *Kant and Stoic Cosmopolitanism*, J. Pol. Phil. (forthcoming 1997) (manuscript at 24, on file with the Columbia Law Review) ("Kant . . . conceive[s] of the passions . . . as natural and precultural . . .").

35. See Meyer, *supra* note 25, at 300 ("I predict: The 'will' has virtually passed out of our scientific psychology today; the 'emotion' is bound to do the same. In 1950 American psychologists will smile at both these terms as curiosities of the past."); see also Martha C. Nussbaum, *Animal Emotions*, in Nussbaum, *Upheavals*, *supra* note 18 (Gifford Lecture II manuscript at 5-6, on file with the Columbia Law Review) ("In their zeal to dismiss the inner world of experience, psychologists in the grip of the newly fashionable behaviorism predicted that emotion would soon disappear from the scientific scene, as a 'vague' and 'unobservable' phenomenon, a relic of our pre-scientific past.").

36. For a highly critical history of these developments, see Lazarus, *supra* note 10, at 3-15.

37. See, e.g., Freud, *Psycho-Analysis*, *supra* note 13, at 22-23 ("[M]ental disturbances are open to therapeutic influence only when they can be identified as secondary effects of some organic disease."); Sigmund Freud, *Project for a Scientific Psychology*, in *The Freud Reader* 86, 87 (Peter Gay ed., 1989) (describing a "psychology that shall be a natural science: that is, to represent psychical processes as quantitatively determinate states of specifiable material particles").

school,<sup>38</sup> which attributed rich cognitive content to the child's emotions of envy, grief, and longing), mechanistic views had and continue to have wide popular influence. Contemporary versions of the mechanistic conception in law and public life frequently show the influence of a multiplicity of sources, including philosophy, psychology, and psychoanalysis.<sup>39</sup>

3. *Arguments Against the Mechanistic Conception.* — The mechanistic conception captures some features of emotional experience, especially the sense we have that emotions are urgent and heated ways of responding to an event, and that they take us over without our consent. On the other hand, it seems quite unable to explain other aspects of emotional experience. If we think, for example, of a person who is experiencing grief, we discover that we cannot describe the emotion without mentioning an object toward which it is directed. The grief of a person who mourns the death of a child cannot adequately be described simply as a force or a current—or even as a force caused by a thought that remains external to the grief—for we must mention that the grief itself is *directed at the child*. If we remove from the experience its character of being focused on that face and form, it becomes something else, a mere attack of high blood pressure or indigestion. Even the felt pain is pain *at the thought of the child's death*<sup>40</sup> and at its distressing consequences. That is what makes it the pain it is; that is what makes it different from the pain of indigestion or muscle cramp.

Furthermore, the object of the emotion is what philosophers generally call an *intentional object*—that is, its role in the emotion depends on its interpretation by the person experiencing the emotion. Suppose the person has been deceived, and her child is in fact alive and well; nonetheless, she feels grief, because, deceived, she sees the child as permanently lost. We often get angry at people for inadequate or mistaken reasons, seeing them as people who have done us some wrong when in fact they have not. Nonetheless, what makes our emotion anger is the way *we* see the person, not the way the person actually is. This “aboutness” is not explained by the mechanistic view.

Finally, emotions embody beliefs, and often very complex beliefs, about their objects—a fact the mechanistic view, once again, fails to explain. Aristotle makes precisely this point, arguing that changes in belief yield changes in emotions. In order to have fear, he argues, a person

---

38. See John Bowlby, *Loss: Sadness and Depression* 273–75, 285–92, 425–39 (1980) (discussing loss and early childhood development); Nancy Chodorow, *The Reproduction of Mothering: Psychoanalysis and the Sociology of Gender* 40–54 (1978) (reviewing object-relations theory and psychoanalytic background); Fairbairn, *supra* note 13, at 162–79 (critiquing Freud and outlining object-relations theory).

39. See, e.g., Thomas C. Grey, *Eros, Civilization and the Burger Court*, 43 *Law & Contemp. Probs.*, Summer 1980, at 83, 90–97 (criticizing some recent Supreme Court jurisprudence on sexual issues and citing decisions and commentators referring to sex as, among other things, a “mysterious . . . force” and an “[i]mpuls[e]”).

40. See Martha Nussbaum, *Emotions as Judgments of Value*, in Nussbaum, *Upheavals*, *supra* note 18 (Gifford Lecture I manuscript at 9–14, on file with Columbia Law Review).

must believe that some impending event threatens his well-being, and that the threat is serious.<sup>41</sup> In order to be angry, he must believe that someone has wronged him (or someone or something dear to him), not inadvertently but deliberately, or perhaps negligently or recklessly, in a more than trivial way. In order to have pity, he must believe that someone else has suffered something seriously bad in a way that he or she does not deserve, or at least deserve fully.<sup>42</sup> If we remove or modify these beliefs, we can expect the emotion to be modified along with them—directly, simply by that change of belief, and not by some further process of behavioral conditioning. If he discovers that not X but Y has done the damage, we can expect his anger to shift from X to Y. If he discovers that the damage never occurred at all, we can expect his anger to go away. If he discovers that the damage was not serious, but trivial, we can expect his anger to become mild irritation. And so forth. Because the mechanistic view treats the belief as external to the emotion and emotion as something that does not respond directly to belief, it will need to invoke some further process of suppression or conditioning to explain the change or diminution of emotion. Mechanists could argue, perhaps, that a change in belief prompts one to engage in some process of suppression or conditioning of the emotion, and that this process yields changes in emotion, but they cannot attribute the change to any inherent connection between emotion and belief. This claim, however, does not accurately represent our experience of changes of this kind. If I am grieving for a loved one in the belief that she has just died, and then I find out that I have been misinformed and she is still living, my grief will go away directly on account of that change in belief; I do not need to undertake any further process of behavioral modification to get rid of it.

For related reasons, the mechanistic view seems to be incapable of explaining the way in which we differentiate emotions. As Aristotle shows, we usually individuate and distinguish emotions with reference to the characteristic beliefs that go with them—in the case of fear, the belief that damage is impending; in the case of anger, the belief that one has been wronged; and in the case of pity, the belief that someone else is incurring serious and undeserved adversity. To defend itself consistently, the mechanistic view apparently must find ways of defining each of these emotions without making reference to these thoughts: for to include the thought inside the definition is to make it a part of the identity of the emotion itself. Can the mechanistic conception do this? It seems doubtful. Fear, pity, and other painful emotions are not clearly distinct in the

---

41. See Aristotelis, *Ars Rhetorica* 1382a, at 86–88 (Rudolfus Kassel ed., 1976) [hereinafter Aristotle, *Rhetoric*] (citations to the *Rhetoric* include the standard “Bekker” section designations followed by the page numbers in the Kassel edition).

42. See *id.* at 1385b–1386b, at 119–23. Aristotle’s is now the dominant view in cognitive psychology. See Lazarus, *supra* note 10, at 12–14 (surveying recent developments in psychology and concluding that “[recent] changes in outlook [have] also brought us back to a kind of ‘folk psychology’ once found in Aristotle’s *Rhetoric*”).

way they *feel*.<sup>43</sup> Even to tell whether a given experience is anger or grief or envy or anxiety often requires inspection of the associated thoughts.

The same is true of the positive emotions—love, joy, gratitude, and hope do not each have a unique feeling-state constantly associated with them. Two emotions may resemble one another in “feel,” and a single emotion, for example love, may encompass a variety of (frequently contradictory) feelings.<sup>44</sup> Of course the mechanist might simply ignore the whole enterprise of defining emotions, and treat the distinctions among them as unimportant. But they are not unimportant features of human experience, and philosophical accounts of emotion commonly acknowledge this.<sup>45</sup> This is one reason why we find no entirely pure example of the mechanistic conception in the history of philosophy. Philosophers are fond of giving definitions, and no known list of definitions of the emotions so much as proposes to define them without reference to belief or thought. Psychologists too have been sufficiently committed to defining the individual emotions that they have viewed the mechanistic conception’s inability to yield adequate definitions as a decisive point against it. For this reason, the experimental results of Schachter and Singer, which showed that subjects differentiated emotions such as joy, anger, and fear not in accordance with their physiological state but in accordance with their beliefs about their situation,<sup>46</sup> were regarded as having sounded the death-knell for the mechanistic conception in cognitive psychology.<sup>47</sup>

Thus, we cannot regard the thoughts or beliefs associated with emotion as simply concomitants or causal prerequisites. If they are needed to identify or define an emotion, and to individuate one emotion from another, this means that they are part of what the emotion itself *is*, constituent parts of its identity.

---

43. This has been shown experimentally, for example in the famous Schachter/Singer experiments, where subjects in whom an identical physiological state had been induced, but who were placed in different situations, identified their emotion differently in accordance with their beliefs about their situation. See Stanley Schachter & Jerome E. Singer, *Cognitive, Social, and Physiological Determinants of Emotional State*, 69 *Psychol. Rev.* 379, 382–93 (1962).

44. There is a further question whether love should be understood as an emotion at all, rather than as a relationship containing emotional components. But even the emotions that are parts of a love relationship are certainly associated with a wide range of feelings.

45. All major philosophical accounts of emotion mentioned in this Article include definitions of distinct emotions. See, e.g., Aristotle, *Rhetoric*, *supra* note 41, at 1378a8–1378a30, at 76–77; Benedict de Spinoza, *Ethics* 137–43 (James Gutman ed. & William H. White trans., 1949); Descartes, *supra* note 13, at 349–52.

46. See Schachter & Singer, *supra* note 43, at 395–98.

47. See James D. Weinrich, *Toward a Sociobiological Theory of the Emotions*, in 1 *Emotion: Theory, Research, and Experience* 129–31 (Robert Plutchik & Henry Kellerman eds., 1980).

## B. *The Evaluative Conception*

1. *The Basic Account.* — If a view of an object, and thoughts about that object, are integral to the experience of emotion, what sort of thoughts must those be?<sup>48</sup> It would appear that the thoughts involved in our examples all have something in common: all appraise or evaluate the object as significant or valuable. We do not get angry over slights we think trivial (or if we do it is usually because this is not really, at some level, what we are thinking). In most cases, anger is associated with the high evaluation of things that matter to us, such as honor, status, the security of our possessions, or the safety and happiness of people we love.<sup>49</sup> It appears that the emotions themselves *contain* an evaluation or appraisal of the object—that is, the appraisal is part of the belief-set in terms of which the emotion will be defined, and these ways of seeing the world are a 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. Sometimes, indeed, the experience of emotion reveals patterns of evaluation of which we might previously have been unaware. A reaction to the loss of a loved one may inform a person about the real importance that loved one had in her life;<sup>50</sup> anger at an insult to a person's appearance may reveal to her that she ascribes more importance to her appearance than she might want to admit; and so forth.<sup>51</sup> The evaluative conception relies on just

---

48. Strictly speaking, one can imagine many forms that a cognitive conception might take, of which the evaluative conception is only one. But it is also the only one that has been prominently defended, and the only one that has real explanatory power, so we shall focus on it in what follows.

49. On the connection between anger and an affront to personal dignity, see generally John G. Milhaven, *Good Anger* (1989).

50. This is one major theme in Marcel Proust's *Remembrance Of Things Past*. See 3 Marcel Proust, *Remembrance Of Things Past* 425 (C.K. Scott Moncrieff et al. trans., 1981) ("A moment before . . . I had believed that this separation . . . was precisely what I wished. . . . But now . . . I felt I could not endure it much longer."); see also Martha C. Nussbaum, *Love's Knowledge* 267 (1990) [hereinafter Nussbaum, *Love's Knowledge*] ("The suffering itself is a piece of self-knowing. In responding to a loss with anguish, we are grasping our love.").

51. See, for example, the summary of this point in Richard Lazarus's book *Emotion and Adaptation*:

When we react with an emotion . . . every fiber of our being is likely to be engaged. . . . The reaction tells us that an important value or goal has been engaged and is being harmed, placed at risk, or advanced. From an emotional reaction we can learn much about what a person has at stake in the encounter with the environment or in life in general, how that person interprets self and world, and how harms, threats, and challenges are coped with. No other concept in psychology is as richly revealing of the way an individual relates to life and to the specifics of the physical and social environment.

Lazarus, *supra* note 10, at 6–7.

this sort of experiential claim when it holds that emotions contain not only thought, but thought of a particular sort, namely *appraisal* or *evaluation*—and, moreover, evaluation that ascribes a reasonably high importance to the object in question.

The value perceived in the object appears to be of a particular sort. It appears to make some reference to the person's own well-being, or to the role of the object in the person's own life. One does not go around fearing any and every catastrophe anywhere in the world, and (so it seems) not even any and every catastrophe one knows to be bad in important ways. What inspires fear is the thought of damages impending that cut to the heart of one's own cherished goals, relationships, and projects. What inspires anger is damage done to someone or something to which one attaches importance in one's own scheme of goals. What inspires revulsion or disgust is an object that seems to threaten to contaminate or damage one's being.<sup>52</sup> We can see, then, that it is no accident that fear, anger, envy, grief, disgust, and the rest should have been grouped together as a family by theorists who agree about little else. For they have a common subject matter: they concern elements of the world (usually elements not fully under the person's control) that are seen to be of vital importance to the person's well-being.<sup>53</sup>

According to the evaluative conception, emotions involve appraisal or evaluation in two distinct ways. First, they contain within themselves an appraisal or evaluation *of* an object. But this also implies that they can

---

52. As we have said, disgust can be of several kinds. Some cases of disgust involve visceral reactions, for example those to certain smells, that may be innate and have little clear cognitive content—although even here the response, no doubt highly adaptive in evolutionary terms, conveys valuable information to the creature, and can thus be said to have an evaluative and cognitive content. See *id.* at 259–60. Most forms of disgust, however, involve learning that has associated the object with danger and contamination. See *id.* at 260.

53. This does not entail that the emotions view these objects simply as tools or instruments of the agent's own satisfaction. See Aristotle, *Ethics*, *supra* note 13, at 1097a–1100a, at 8–16. They may be invested with intrinsic worth. One way of understanding this is to think of the ancient Greek notion of *eudaimonia* or human flourishing. According to this idea, people are always pursuing the integrated realization of their own system of goals and ends, even though this does not at all imply that the various ends are simply means to one's own feelings of satisfaction or happiness. See J.L. Austin, Agathon and Eudaimonia in the *Ethics* of Aristotle, reprinted in *Philosophical Papers* 1, 12–31 (J.O. Urmson & G.J. Warnock eds., 3d ed. 1979). A friendship may be pursued as good in itself; nonetheless, what makes it important for a person is that it is *hers*, a part of her own scheme of ends and goals. See Bernard Williams, *Egoism and Altruism*, in *Problems of the Self: Philosophical Papers 1956–1972*, at 250, 250–65 (1973) (noting that the possible results of even purely altruistic desires “just came[ ] down to a question whether [the *individual's* desire] is satisfied or not”). Pity might appear to be less “*eudaimonistic*” than the other emotions: but as Adam Smith argues well, a person will have pity for a disaster befalling another only to the extent that she has managed to move that other person close to her in imagination, making that person part of her own scheme of what matters and is worth pursuing. See Adam Smith, *The Theory of Moral Sentiments* 140–44 (D.D. Raphael & A.L. Macfie eds., Liberty Classics 1982) (1759).

themselves *be* evaluated. For people ascribe significance to parts of the world in many different ways, and these thoughts may be either correct or incorrect, and (a separate point) formed in a reasonable or unreasonable manner. People's emotions may contain mistakes in two different areas. Sometimes emotions are inappropriate because the person was simply wrong about what had happened, or about who was involved. (Sometimes we may want to blame the person for that error and sometimes not.) For example, a person might become afraid because of the incorrect belief that the very presence of an African-American man on the other side of the street constituted a threat of rape. In this case, we will probably want to blame her for holding the incorrect factual belief. Sometimes, by contrast, our criticism of an emotion will focus on the value-appraisals themselves, or, we might say (to show we are not making a crude distinction between fact and value), the value-facts. For example, we may say that a person's anger is inappropriate if it ascribes overwhelming importance to the fact that someone has forgotten the person's name (an example given by Aristotle in the *Rhetoric*, showing us the constancy of human vanity over the centuries).<sup>54</sup> Someone provoked to an intense reaction by such a trivial event would be thought irrational and would be criticized. On the other hand, intense anger at the murder of one's own child, or at an assault that violates one's own bodily integrity, seems a perfectly appropriate assertion of love or dignity.<sup>55</sup>

If we claim that emotions involve evaluative thought, we naturally begin to ask questions about the sort of evaluations reasonable people ought to make. Thinking this way, Aristotle holds that the virtuous person observes the mean (by which he means not a middle course, but a course of appropriateness)<sup>56</sup> with regard to both action and emotion, and that the criterion of that appropriateness should be found by asking what a person of practical wisdom would do and feel in the situation.<sup>57</sup> The person of practical wisdom is an ideally reasonable agent who has a well-formed character and who embodies the "reputable views"<sup>58</sup> of the

---

54. See Aristotle, *Rhetoric*, *supra* note 41, at 1379b30–1379b35, at 81.

55. As Aristotle says, "The person who lets himself and his loved ones get trampled on and overlooks it seems like a slave." Aristotle, *Ethics*, *supra* note 13, at 1126a7–1126a8, at 81. Recall, in our own time, the widespread public criticism of Michael Dukakis because he did not display anger at the mere prospect of his wife's rape. See, e.g., Mary McGrory, *Deadly Seriousness*, *Wash. Post*, Nov. 10, 1988, at A38 (describing Dukakis's response as "inadmissibly impersonal"); see also *The Presidential Debate: Transcript of the Second Debate Between Bush and Dukakis*, *N.Y. Times*, Oct. 14, 1988, at A14. In value mistakes as well, incorrectness should be distinguished from unreasonableness. Some beliefs that are in fact incorrect might be reasonable to hold, given a certain state of scientific or social knowledge.

56. For an extensive discussion, see J.O. Urmson, *Aristotle's Doctrine of the Mean*, in *Essays on Aristotle's Ethics* 157 (Amélie O. Rorty ed., 1980).

57. See Aristotle, *Ethics*, *supra* note 13, at 1105b, at 29–30; see also Urmson, *supra* note 56, at 157–59.

58. This is the best translation of *endoxa*, on which Aristotle bases his normative ethical arguments.

community at their best, after these have been scrutinized and sifted by critical argument.<sup>59</sup> In other words, he embodies a recognizable ideal of appropriate behavior and evaluation, although in many respects this may be distinct from average or common behavior and evaluation.<sup>60</sup> In a related manner, Adam Smith evaluates the emotions by asking what the emotions of his ideally reasonable person, the "impartial spectator," would be.<sup>61</sup> Aristotle's project and Smith's differ in detail, but they embody a common strategy through which partisans of the evaluative conception, in philosophy, in public life generally, and in the law, have sought to make the evaluations of people's emotions that this conception invites.

We should insist at this point on the distinction between having the emotion and acting in accordance with it. A person who becomes violently angry may or may not judge it appropriate to express that anger in a violent action. She might, for example, believe that the law should handle such matters. Or she might be a pacifist who holds that it is always wrong to act violently toward another human being.<sup>62</sup> One may be afraid without even desiring to run away—as Aristotle insists, describing the courageous person who nonetheless will fear the loss of his life. In such cases, the emotion, if present, will still contain an appraisal of the situation as a bad one: the angry pacifist still thinks the injustice is very bad, and the fearful but constant soldier still thinks that the loss of his life is very bad. Their actions, however, are not determined by that one judgment alone, but by a whole host of judgments. In other words, once one has granted that a given emotional response is appropriate, one needs to ask many further questions before one can determine that a given related action is appropriate—and conversely.

Now that we have specified precisely the type of appraisal or evaluation involved in emotion, we can see that the evaluative conception can explain the intuitive phenomena that the mechanistic conception claims on its side. We identified those phenomena as the link between emotion and passivity, the sense that emotion is not really part of one's self, and the sense of urgency we ascribe to emotional experience.<sup>63</sup> According to the evaluative conception, we feel passive in emotion because in emotion we recognize that some object or objects outside of us have great signifi-

---

59. See David Wiggins, *Deliberation and Practical Reason*, in *Essays on Aristotle's Ethics*, *supra* note 56, at 221, 234–37.

60. His reasonableness has both a substantive and a procedural component: he is envisaged as someone who gets the right beliefs not just by accident, but as the result of good critical reflection. See *id.*

61. See Smith, *supra* note 53, at 26–27, 69–70.

62. One may also choose a violent action without the emotion—for example, because ordered to perform it by one's commanding officer.

63. See *supra* text accompanying notes 28–29.



cance for our well-being, whether for good or for ill.<sup>64</sup> In grief, for example, I feel buffeted by the world because I am recognizing that I have just lost someone who is important to my flourishing. In emotion we recognize our passivity before the ungoverned events of life, with respect to our most important goals and projects. Of course, since this recognition has its felt side as well, we will also feel passivity toward this feeling, being seized by tears, or laughter, and so forth. But this passivity results from our passivity to the events that occasion the emotional response.

Emotions feel like external forces distinct from the self because, once again, emotions register transactions with a world outside of ourselves about which we care deeply, in which we have invested a great part of our own well-being. Selves in that sense incorporate parts of the world into themselves, whenever they ascribe importance to such external and unreliable items as friends, children, citizenship, or a country. This means that in unhappy times people will sometimes report feeling torn limb from limb, while in happier times they may feel filled with a marvelous sense of integration or wholeness.<sup>65</sup>

Finally, this view can explain why the emotions seem to have urgency or heat: because they concern our most important goals and projects, the most urgent transactions we have with our world. This view in fact seems to explain urgency better than does the mechanistic conception. For if there is urgency in being hit by a gust of wind, it is not after all an unthinking sort of urgency. The urgency, if it is there, resides not in the wind itself, but in my thought that my well-being is threatened by that wind. The evaluative conception, bringing thought about well-being right into the structure of emotion, shows why it is the *emotion itself*, and not some additional reaction to it, that has urgency and heat.

2. *History and Development.* — Like mechanistic theorists, evaluative theorists can trace their view to Plato's *Republic*—in this case, to his position that the emotional part of the soul is an "ally of belief" and responds to the socially shaped formation of belief.<sup>66</sup> Plato goes so far as to argue that certain emotions he finds pernicious, such as fear and pity, can be pretty well eliminated from life by the social management of evaluative beliefs.<sup>67</sup> Aristotle's discussion of emotion in the *Rhetoric*<sup>68</sup> and the *Nicomachean Ethics*,<sup>69</sup> of particular importance for the history of popular

---

64. This recognition need not be conscious: frequently, as in one's fear of one's own death, the recognition that an impending event is bad may guide one's actions and reactions without one being fully conscious of it.

65. See James R. Averill, *Anger and Aggression: An Essay on Emotion* 185–91 (1982).

66. See Plato, *supra* note 26, at 440–41.

67. See *id.* at 386–88.

68. See Aristotle, *Rhetoric*, *supra* note 41, at 1378a20–1378a25, at 76–77 ("The emotions are all those feelings that so change men as to affect their judgments, and that are also attended by pain or pleasure.").

69. See Aristotle, *Ethics*, *supra* note 13, at 1125b26–1126b10, at 80–82 ("The man who is angry at the right things and with the right people, and, further, as he ought, when he ought, and as long as he ought, is praised.").

and legal thought, stresses the role of evaluative belief and judgment.<sup>70</sup> Emotions are held to be an assessable part of a person's character.<sup>71</sup> Aristotle argues that emotions are very often appropriate to the situations in which people find themselves: the "person of practical wisdom," his norm of the reasonable person, will fear his own death,<sup>72</sup> grieve at the deaths of loved ones,<sup>73</sup> pity the calamities that befall an undeserving person,<sup>74</sup> and get angry at serious wrongs done to himself or his loved ones.<sup>75</sup> On the other hand, Aristotle considers many emotions people actually experience to be inappropriate because they embody false evaluations—for example, an excessive attachment to status<sup>76</sup> or money.<sup>77</sup> His conception (developed also in a very influential form by Thomas Aquinas)<sup>78</sup> shows us how an evaluative conception may be based on a careful analysis of ordinary beliefs and yet end up being very critical of many ordinary practices.<sup>79</sup>

The Greek Stoics who lived shortly after Aristotle developed the evaluative conception in an especially detailed and compelling manner, presenting most of the arguments that we have given here about its structure and its explanatory power.<sup>80</sup> Through their own writings and those of later Roman thinkers such as Seneca<sup>81</sup> and Cicero,<sup>82</sup> the Stoics exerted a major influence on all subsequent philosophical theories of emotion,

---

70. See, e.g., Nussbaum, *Therapy*, supra note 15, at 80 (describing how, in Aristotle's view, "[e]motions have a very intimate relationship to beliefs, and can be modified by a modification of belief").

71. Of particular interest is the account of the virtue of "mildness of temper" in *Ethics* IV.5, where he describes the proper balance where anger and provocation are concerned. See Aristotle, *Ethics*, supra note 13, at 1125b26–1125b35, at 80.

72. See id. at 1115a6–1115a29, at 53–54.

73. See id. at 1169a3–1170a19, at 193–96; Aristotle, *Rhetoric*, supra note 41, at 1380b34–1381a11, at 84.

74. See Aristotle, *Rhetoric*, supra note 41, at 1385b11–1386a3, at 95–96.

75. See Aristotle, *Ethics*, supra note 13, at 1125b31–1126a8, at 80–81.

76. See id. at 1125b1–1125b11, at 79–80.

77. See id. at 1119b22–1121b10, 1128b4–1128b35, at 69, 87–88.

78. See, e.g., Thomas Aquinas, *Summa Theologica*, in 1 *Basic Writings of Saint Thomas Aquinas* Q.81, art.3, at 774–76 (Anton C. Pegis ed., 1945) (citations to *Summa Theologica* include the standard question and answer designations followed by the page numbers in the Pegis edition) ("[T]he sensitive appetite is naturally moved, not only . . . in man by the cogitative power which the universal reason guides, but also by the imagination and the sense.").

79. See, e.g., Aristotle, *Ethics*, supra note 13, at 1121a30–1121b1, 1128b4–1128b35, at 69, 87–88.

80. See, e.g., 2 *The Hellenistic Philosophers* 404–18 (A.A. Long & D.N. Sedley eds., 1987) (collecting Stoic commentaries on emotions).

81. See, e.g., Seneca, *On Anger*, supra note 15, at 3.13.1–6, at 89–90; Seneca, *On Mercy*, in *Seneca: Moral and Political Essays*, supra note 15, at 2.3–7, 160–64 [hereinafter *On Mercy*] (citations to *On Mercy* include the standard section designations followed by the page numbers in the Cooper edition).

82. See, e.g., Cicero, *Tusculan Disputations* IV.iv.8–vi.14, at 337–43 (J.E. King trans., 1927) (citations to the *Tusculan Disputations* include the standard section designations followed by the page numbers in the King edition).

particularly those of Spinoza,<sup>83</sup> Rousseau,<sup>84</sup> and Adam Smith.<sup>85</sup> The subsequent history of the evaluative conception is largely the story of the reception of Aristotelianism and Stoicism.

The evaluative conception is once again the dominant account of emotion in contemporary philosophy and psychology. In philosophy, the turning away from a once-popular mechanism and behaviorism was inspired by a revival of Aristotelianism,<sup>86</sup> and also by some cryptic comments of Ludwig Wittgenstein<sup>87</sup> that were developed by Anthony Kenny into a devastating critique in the important book *Action, Emotion, and Will*.<sup>88</sup> More recently, a variety of thinkers with subtly different points of view have defended the evaluative view.<sup>89</sup> The current revival of interest in ancient Greek accounts of character and virtue has also enriched the argument in favor of this conception.<sup>90</sup>

Similar developments took place somewhat later in cognitive psychology. Empirical results led psychologists to the conclusion that the subject's own evaluations of her situation were an essential part of her emotions, and of what differentiated one emotion from another.<sup>91</sup> Evo-

83. See, e.g., Spinoza, *supra* note 45, at 127–252.

84. See Jean-Jacques Rousseau, *Emile or On Education* 211–50 (Allan Bloom ed. & trans., Basic Books 1979) (1762).

85. See D.D. Raphael & A.L. Macfie, Introduction to Smith, *supra* note 53, at 5–10 (“Stoic philosophy is the primary influence on Smith’s ethical thought.”); see also *id.* at 265–342 (citing heavily to Stoic philosophers and their predecessors); *id.* at 34–38 (reflecting Stoic views in his discussion of the “unsocial Passions”). It is important to distinguish the Stoics’ *analysis* of emotions (their influential defense of the evaluative over the mechanistic conception), see *supra* notes 80–82 and accompanying text, from their controversial *normative thesis*, according to which all the evaluations involved in the major emotions are mistaken, because it is always mistaken to ascribe any great importance to things outside of oneself that one’s own will does not control. See Smith, *supra* note 53, at 275 n.k. (criticizing the Stoic normative thesis). This extreme position about value entails that emotions are never parts of a good character: there are no good reasons to get angry, to be afraid, and so forth, since the only really important things, one’s own reason and will, can never be damaged. For a Stoic, then, there would be no such thing as reasonable provocation: Stoic writings delight in stories of people who meet the maximum provocation and insult with a cool unruffled demeanor.

86. See Lazarus, *supra* note 10, at 13–14.

87. See, e.g., Ludwig Wittgenstein, Zettel 101e, ¶ 576 (G.E.M. Anscombe & G.H. von Wright eds. & G.E.M. Anscombe trans., 1967) (“For the interesting thing is not that I do not infer my emotion from my expression of emotion, but that I also do not infer my later behavior from that expression, as other people do, who observe me.”); *id.* at 87e, ¶ 493 (“But these sensations are not the emotions.”).

88. See Kenny, *supra* note 32, at 29–51 (reviewing self-contradictory mechanistic theories and research and concluding, at 49, that “[t]he occasion on which an emotion is elicited is part of the criterion for the nature of the emotion”).

89. See, e.g., Ronald de Sousa, *The Rationality of Emotion* at xv–xvi (1987); William Lyons, *Emotion* (1980); Robert C. Solomon, *The Passions: Emotions and the Meaning of Life* (1993).

90. On this revival, see Martha C. Nussbaum, *Virtue Revived*, *Times Literary Supplement* (London), July 3, 1992, at 9.

91. See generally Lazarus, *supra* note 10, ch. 4 (discussing research on cognition and emotion). See also Keith Oatley, *Best Laid Schemes: The Psychology of Emotions* 44–68

lutionary biology has contributed to these developments by showing that emotions frequently provide valuable information to an individual, and by suggesting that emotions evolved precisely because of their information-providing function.<sup>92</sup> Meanwhile, in psychoanalysis, Freud's mechanistic conception almost immediately encountered criticism from theorists of the "object-relations" school, who argued that the infant's early life contains complicated relationships to external objects and persons, and emotions that are based on the infant's thought about the all-important contribution of those objects.<sup>93</sup> Today such conjectures about the rich cognitive life of infancy are receiving increasing experimental support.<sup>94</sup>

Anthropology has also made valuable contributions to our understanding of emotions, providing much evidence of the role played in emotion by social norms that shape evaluation of objects.<sup>95</sup> Detailed anthropological analyses of the emotion-conceptions of a given society show us vividly how emotions actually differ from impulses, and how much evaluative thinking they involve.<sup>96</sup> This work has been pivotal in showing the extent to which our emotional life is not "natural" or pre-social at all, as the mechanistic view typically holds, but thoroughly bound up with socially learned patterns of evaluation.

One further clarification can now be made. This history shows that a person who calls emotions "irrational" may mean one of two very different things. One would be that the mechanistic conception is true and the evaluative conception false: emotions have nothing (or nothing much) to do with reasoning or thought. Another, however, would be that emotions are bound up with thoughts that are confused, unreliable, or sub-par. One could say this without advocating the mechanistic con-

---

(1992) (reviewing literature and research supporting a "communicative theory" of emotions focused on goals); Andrew Ortony et al., *The Cognitive Structure of Emotions* ch. 2 (1988) (reviewing theories of emotion and presenting their theory that different emotion types result from differences between the circumstances in which they generally occur); Seligman, *supra* note 14, ch. 3 (reviewing studies showing relationships between individuals' evaluations of their situations and learned helplessness).

92. See de Sousa, *supra* note 89, at 43-44, 47-106 (analyzing biological and teleological models of emotional evolution and development); Lazarus, *supra* note 10, at 50-53, 68-69 (acknowledging relevance of evolutionary biology).

93. See Bowlby, *supra* note 38, at 9-22 (summarizing scholarly controversy surrounding mourning in infants, and reviewing studies and literature supporting mourning among the very young).

94. For a sophisticated examination of the cognitive life of a baby from six weeks to four years old, see Daniel N. Stern, *Diary of a Baby* (1990); see also Daniel N. Stern, *The First Relationship: Mother and Infant* chs. 3, 7 (1977) (discussing how infant and mother/care-giver interactions function and result in the development in the infant of interest, delight, and boredom).

95. For especially valuable examples, see Jean L. Briggs, *Never in Anger: Portrait of an Eskimo Family* (1970) (chronicling the emotional life of an Eskimo community); Lutz, *supra* note 24, at chs. 1-2 (discussing the cultural construction of emotions, and describing her anthropological fieldwork on emotions in Micronesia).

96. See, e.g., Lutz, *supra* note 24, ch. 6 (analyzing the emotion of "justifiable anger").

ception. In fact, it is hard to see how someone could endorse this view without espousing the evaluative conception. The leading proponents of this claim historically, the Stoics and Spinoza, have been strong evaluative theorists, who think that people ascribe much more importance to honor, money, love, and so forth, than they ought to. People frequently confuse the two senses of "irrational" in informal usage;<sup>97</sup> sometimes they use Stoic normative language and then misread it, taking it to imply a mechanistic conception. For these reasons and others, unclear contrasts between "emotion" and "reason" abound in our talk about emotions.

3. *Different Species of the Evaluative Conception.* — Several different versions of the evaluative conception can be distinguished, in accordance with the role or roles they ascribe to belief in the emotion. We may distinguish these versions by identifying four theses, which we shall call the necessary condition thesis, the constituent part thesis, the sufficient condition thesis, and the identity thesis. All versions of the evaluative conception agree that certain cognitive attitudes (usually, beliefs or judgments) are *necessary conditions* for an emotion: in order for grief to be present, for example, the person must believe that she has suffered a serious loss. Most versions of the conception understand that necessary role as internal: the beliefs are necessary as *constituent parts* of the emotion, not as external causes of something that does not in its nature contain belief.<sup>98</sup> The reason for putting the belief right into the emotion as a constituent part is, as we have already argued, that it appears impossible to give a fully adequate definition of what a given emotion is, or to say how it differs from other emotions, without mentioning the beliefs involved. To separate pity from fear, fear from grief, we cannot rely on the felt quality of the pain alone; only an inspection of the characteristic thought patterns allows us to discriminate. Thus, in defining pity, and distinguishing it from fear and grief, we need to include in the definition not only the fact that it is a painful emotion, but also the fact that it is directed at the suffering of another person, and that it involves the belief that the suffering is serious, that the person did not fully deserve it, and so forth. The arguments for the necessary condition thesis and the constituent part thesis seem very strong. Hereafter, in discussing the evaluative conception, we shall assume that these two features are part of it: the cognitive attitudes in question are necessary for the emotion, and necessary as constituent parts, whether or not we hold that there are also other constituent parts.

Partisans of the evaluative view have, however, made two further claims. The first is that the relevant beliefs or other cognitive attitudes

---

97. This kind of confusion already occurred in antiquity: Galen, attacking the Greek Stoic thesis, tries to convict them of inconsistency for saying both that emotions are judgments and that they are *alogoi*, "irrational." See Galen, *supra* note 21, at 239, 253–55.

98. See Nussbaum, *Therapy*, *supra* note 15, at 80.

are actually *sufficient conditions* for the emotion in question.<sup>99</sup> For example, if a person really believes she has suffered a serious loss, that is sufficient for grief; if we don't find grief there, we have reason to doubt that the person has the beliefs we have supposed. Aristotle appears to rely on a thesis of this kind when he writes in the *Rhetoric* that an orator can actually *produce* certain emotions in his audience by causing them to have certain beliefs, or calm them down by taking certain beliefs away.<sup>100</sup> Many good arguments support this thesis. One can accept it whether or not one thinks that the emotion itself has further affective aspects that are distinct from the beliefs or cognitions: for one may simply hold, as Aristotle apparently does, that the presence of the beliefs suffices to produce these other elements as well. The feelings of pleasure and pain are pleasure or pain "in" or "at" what one believes to be the case. Given the evaluative content of the belief, which connects it with one's own well-being, having such a belief will prove sufficient to bring about the pleasure, or the pain.

As for apparent counterexamples to the thesis, in which a person seems to have all the relevant beliefs but lacks the related emotion, these usually prove, on further inspection, to be cases in which the person does not really have all the relevant beliefs, including the evaluative ones. One may of course believe that someone has died without feeling grief. What seems doubtful is whether one can believe that a person whom one sees as absolutely central to one's life, to whom one is deeply attached, has died, without having grief. One may temporarily fail to experience such grief—a commonly discussed phenomenon in the literature on mourning, which is generally explained with reference to a failure to internalize cognitively the magnitude of the loss that has occurred.<sup>101</sup> But when one does recognize that one's life has suffered such a devastating blow (in a case where one really does depend centrally on the person who has died), grief, it is held, will result. The sufficient condition thesis is perhaps a more controversial element in the evaluative conception than the claims of necessity and of constituent parthood. But it appears to be well-grounded, and it sheds light on the ways in which our beliefs about events, not to mention public rhetoric, provoke an emotional response.

---

99. See *id.* at 371–72 (exploring the function of belief and listing sources). Notice that if one holds this thesis without the constituent-part thesis, one may not have an evaluative conception of what the emotion itself is: for one might then just hold that the belief is a sufficient condition of emotion as an external cause of something that in its nature is not at all belief-like. The Stoic thinker Zeno, for example, seems to have said that certain beliefs necessarily produce a fluttering feeling, and that this fluttering feeling was what the emotion actually is. See *id.* at 372.

100. See, e.g., Aristotle, *Rhetoric*, *supra* note 41, at 1380a2–1380a4, 1380b29–1380b33, at 81, 84.

101. See Bowlby, *supra* note 38, at 116 (describing the "Phase of Numbing"); see also 3 Proust, *supra* note 50, at 546 (reflecting on confused and competing images of his dead lover).

Finally, the strongest and most controversial thesis of all is the *identity* thesis: that is, Chrysippus's claim that emotions are simply identical with certain value-laden cognitive attitudes.<sup>102</sup> Grief, for example, just is a certain type of judgment about the seriousness of a loss in relation to the person's well-being. This claim initially seems paradoxical, for we tend to think that bodily sensations and psychological feeling-states of many kinds are involved in emotion, and that these are not themselves parts of a judgment or belief. Our overall argument does not rely on this strong thesis, and we shall therefore not go into detail about the arguments in its favor, though we think them powerful.<sup>103</sup> Briefly put, however, the identity thesis derives its support from the fact that no concretely specifiable body-state or feeling-state can be identified as absolutely necessary for a given emotion—for example, grief. There usually will be some such elements present, but if a given one or ones fail to be present (if the mourner's blood pressure is low, or her eyes free from tears), we do not usually withdraw our ascription of grief if we are satisfied that the cognitive elements of grief are present. What we do look for, and insist on finding, is an awareness of a loss that is seen as extremely serious, with significant impact on the mourner's own life. Only the absence of this value-laden awareness would cause us to withdraw an ascription of grief; this fact, in turn, suggests that only that set of cognitive attitudes belongs in the definition of the emotion. For this reason the identity thesis looks plausible and worth defending; we shall not, however, defend it further here.

In what follows, then, we shall assume that what we call the evaluative conception is committed both to the necessary condition and to the constituent part theses, and, most of the time anyway, to the sufficient condition thesis. The sufficient condition thesis will lead us to speak of "emotions" in cases (for example, of duress or self-defense) where some readers may not always feel intuitively convinced that an emotion need be present. This does not affect our argument. Our position rests on the modest claim that factual and evaluative judgments form at least a part of emotions such as fear or anger. We believe that the acceptance of those judgments is sufficient for the emotion, but if the reader does not, she may simply regard us as speaking about the cognitive *components* of the emotion, and their importance for legal doctrine. The identity thesis we may leave to one side.<sup>104</sup>

---

102. See Nussbaum, *Therapy*, *supra* note 15, at 366–72.

103. See Nussbaum, *Upheavals*, *supra* note 18; Solomon, *supra* note 89, at 125–27.

104. As for terms such as "belief" and "judgment": the evaluative conception, as it has been developed in recent writing, makes room for a number of different types of cognitive attitudes, not all involving a grasp of linguistically formulable propositions; all, however, involve appraisal of an object. See Lazarus, *supra* note 10, at 107–12; Nussbaum, *Upheavals*, *supra* note 18 (Gifford Lecture II manuscript at 28–34, on file with the Columbia Law Review); see also work on animal and infant psychology in Lazarus, *supra* note 10, at 117–18, 180–84; and Oatley, *supra* note 91, at 88, 93–107, 191–94. For

4. *The Role of Social Factors.* — An evaluative theorist need not deny that emotions are part of our evolutionary heritage and have an inherited biological basis.<sup>105</sup> It seems difficult to deny, however, that this inherited material is shaped by society.<sup>106</sup> The social group plays two distinct and complementary roles: it is the source of the child's evaluative learning; and it shapes what is to be learned, defining what counts as good and valuable in ways that vary to some extent from one society to another.

Cultural variation in emotion has a number of different dimensions. First, societies construct norms for the proper expression of emotion in behavior, even when the underlying emotion is basically the same.<sup>107</sup> For example, although the experience of romantic love is probably very similar in England and in the United States, norms for its public expression vary considerably. Second, societies construct norms concerning the appropriateness of whole emotion-types; and these social norms affect, it would seem, not only the expression of emotion, but also the experience of emotion itself. Thus a society that teaches that sexual love is always in some degree sinful constructs an experience of sexual love that will differ from that one might have in a society that had no such teaching: love itself will be experienced in close connection with guilt and shame.<sup>108</sup> Third, societies may contain specific types of emotion that are not found in other societies—either because they focus on entities not known in those societies (a specific Scandinavian type of fear associated with the forest will be absent in Bangladesh), or because societies have simply patterned things differently as a result of their specific history (courtly love overlaps with, but is not the same as, modern American romantic love).<sup>109</sup>

A single society will also, especially in the contemporary world, contain a plurality of emotion-systems. These systems will differ with respect to each of the three sources of variation described above. In contempo-

---

additional discussion of appraisal of objects and infant cognition, see Ortony, *supra* note 91, at 28, and Stern, *Diary of a Baby*, *supra* note 94, at 47–71.

105. See Lazarus, *supra* note 10, at 51–52; Oatley, *supra* note 91, at 140–48; Ortony, *supra* note 91, at 26–33.

106. See Martha C. Nussbaum, *Constructing Love, Desire, and Care*, in *Laws and Nature* (David Estlund & Martha C. Nussbaum eds., forthcoming 1996).

107. See, e.g., Lutz, *supra* note 24, at 40–47 (surveying research approaches); Paul Heelas, *Emotion Talk Across Cultures*, in *The Social Construction of Emotions* 234, 234–61 (Rom Harré ed., 1986) (discussing studies from all over the world, and concluding the “[e]motional elements” might be universal; “emotional experiences” are not); H. Morsbach & W.J. Tyler, *A Japanese Emotion: Amae*, in *The Social Construction of Emotions*, *supra*, 289, 289–305 (contrasting Japanese and American norms).

108. For variation in anger, see Seneca's description of Roman mores in his time in Seneca, *On Anger*, *supra* note 15, at 2.35.3–36.6, at 73–75. See also *id.* at 3.18.1–21.5, at 94–98 (listing prominent Romans who exhibited inappropriate and outrageous anger). In addition to the ideal society depicted by Seneca and other Stoics, see generally Briggs, *supra* note 95 (describing relationships, emotions, and evaluative mores in an Eskimo community).

109. See Staten, *supra* note 16, at 106–07.



rary America, we find many different norms of emotional expression, many different ideas about the appropriateness of emotion-types such as sexual love and anger, and many different concrete varieties of emotion.<sup>110</sup> Probably even most individuals in America are sites of competition among different norms in these three areas. One and the same citizen may grow up with strict norms of emotional reticence, and then learn from the sexual revolution and other social changes to value a certain sort of public display.<sup>111</sup> He or she may be swayed, at different times, both by guilt about sexual love and by the appeal of attitudes that assail guilt. He or she may be able not only to understand intellectually, but also to respond to and enact, the very different patterns and experiences of love suggested, for example, by the Beatles, Wagner, and Billie Holiday.

These facts of social variation in emotion provide further grounds for preferring the evaluative conception to the mechanistic conception; it can accommodate them well, whereas the mechanistic conception cannot. But the fact of social variation creates some delicate problems when, having understood the evaluations that are internal to a given person's emotions, we then turn to the task of evaluating those evaluations for their appropriateness or reasonableness. The facts of social variation warn us that we must ask, "Whose ideas of reasonableness?" When we are asking about anger, for example, do we look for our norm of reasonableness to the Utku Eskimos<sup>112</sup> or to the ancient Romans,<sup>113</sup> or to some critical norm that transcends both cultures?<sup>114</sup> If the answer is that we should look to our fellow-citizens and define the emotion and its norms as they do—by no means an obvious answer, since we might have good reasons to think some prevailing norms unreasonable<sup>115</sup>—we still must ask, which fellow citizens, and why those?

### C. *Emotions, Character, and Moral Education*

Understanding the nature of emotion has practical, not merely theoretical, value. Emotions motivate behavior. Accordingly, if we have an interest in affecting the behavior of another person—whether a child, a student, or a fellow citizen—we should also take an interest in that per-

---

110. See, e.g., Erich Segal, *Love Story* 45–51, 58–66 (1970) (contrasting the different outlooks on romance in families of different ethnicities).

111. See *infra* text accompanying notes 368–372 for a changing conception of appropriate anger.

112. See Briggs, *supra* note 95, at 328–37.

113. See, e.g., Seneca, *On Anger*, *supra* note 15, at 1.5.1–6.5, at 23–25 (examining whether anger is natural).

114. Note that, depending on the way in which we answer this question, we may be in a position to declare an entire local emotion-category unreasonable. If, for example, we discover that anger in Rome is generally linked with ideals of competitive self-assertion over status and power, and if we should judge those norms to be in some respects unreasonable, we might judge that Roman anger was in a general way unreasonable.

115. See, e.g., *infra* text accompanying notes 349–357.

son's emotional life. The evaluative and mechanistic views support radically different programs for educating emotions.

The mechanistic view focuses on noncognitive procedures, such as conditioning, behavior modification, and suppression.<sup>116</sup> It pictures the angry or hateful person as a seething cauldron of inappropriate impulses with a lid held tightly on; education strengthens the hand that holds the lid down. To Kant, for example, the goal of an education in virtue is a strengthening of the will, so that it can resist natural inclinations; virtue is conceived as a type of strength, in which respect for law overcomes feeling and inclination.<sup>117</sup> But because emotions are viewed as mere impulses or drives, it makes little sense on the mechanistic view to speak of shaping or reforming the content of a person's emotions themselves. The most we can aspire to is a noncognitive regimen that trains individuals to experience or fail to experience emotions in different situations, in much the same way that Pavlov's dogs were trained to salivate upon the toll of a bell.<sup>118</sup> But most versions of the mechanistic view hold out little hope that conditioning will render the emotions governable: thus they always focus on the need to cultivate mechanisms of suppression and indulgence as the primary devices to control emotion.

The evaluative view, in contrast, considers emotional education to be closely bound up with moral education.<sup>119</sup> Because cognitive appraisals are integral to emotions, the educator of emotions must address herself to her pupil's beliefs, especially about matters of value.<sup>120</sup> Thus, a child

---

116. See, e.g., B.F. Skinner, *Beyond Freedom and Dignity* 27-43 (1971) (describing "operant conditioning" and manipulation of "behavioral processes characteristic of the human organism").

117. See, e.g., Immanuel Kant, *The Metaphysics of Morals* Part II: *The Metaphysical Principles of Virtue*, in *Kant's Ethical Philosophy*, *supra* note 20, at 53: "Virtue is the strength of man's maxim in obeying his duty. All strength is known only by the obstacles it can overcome; and in the case of virtue the obstacles are the natural inclinations, which can come into conflict with moral purpose." The commandment of virtue is "that he should bring all his capacities and inclinations under his authority (that of reason)." *Id.* at 67. For Kant, this is equivalent to rendering the inclinations non-influential in moral choice, giving them no significant speaking role at all, a state that he describes as "Apathie," or passionlessness. The "duty of apathy" is "the prohibition that man should not let himself be governed by his feelings and inclinations." *Id.* at 68. "The true strength of virtue is the mind at rest." *Id.* To the claim that virtue could be enhanced by emotional excitation Kant replies that this is the mere "apparent strength of a fever patient," a "glittering appearance which leaves one languid." *Id.*

118. Thus Kant concedes that it may sometimes be effective to condition ourselves to feel pity for the sick by going around to visit sickrooms and hospitals, thus training that inclination and giving it strength; on occasion, he argues, pity, though a sensuous inclination and not a judgment, may prove an efficacious supplement to the judgment of duty. See *id.* at 122.

119. See Nancy Sherman, *The Fabric of Character: Aristotle's Theory of Virtue* 2, 49-50, 162, 165-71 (1989); see also Aristotle, *Ethics*, *supra* note 13, at 1114b, at 51-52; Seneca, *On Anger*, *supra* note 15, at 2.20.1-21.11, at 58-61 (emphasizing the role of the mind in physical manifestations of emotion).

120. See Sherman, *supra* note 119, at 157-60.

who gets angry too often, or at the wrong occasions, will be approached not with noncognitive techniques of habituation, but with reasoning and instruction. Some things, the parent will say, are not worth getting angry about. Or: you should not resent the attention the teacher is paying to other children, since it is fair and good for other children to have their share. Or: you should not fear someone whose skin color is different from your own, since that difference is simply not threatening. Or: when strangers ask you to go somewhere for an ice cream, you should view that as ominous, and an occasion for fear. In these countless ways, parents and other educators shape the cognitive content of children's emotions—by teaching them how to appraise the world around them.

The difference between the evaluative and mechanistic programs touches on a number of practical issues. One is whether we ought to make moral assessments at all about a person's emotional life or *character*. The mechanistic view suggests that we should do so only in a limited way. Emotions, on this account, are mere drives or impulses that do not contain beliefs or appraisals. For this reason, it makes about as much sense to blame a person for being inappropriately fearful, angry, or hateful as it does to blame her for having a bad case of heartburn. If a person *behaves* inappropriately while under the influence of such an emotion, we might say that her impulse was too strong and that she should have constrained or suppressed it better; but we cannot say that her emotion, by itself, embodied a morally inappropriate way of seeing the world that should have been corrected through moral education. Partisans of the evaluative conception, in contrast, hold that individuals should in general be held responsible for their character, including its emotional elements, because in general it is up to individuals to shape their own evaluations of people and things in accordance with good norms.<sup>121</sup> When they encounter someone who gets angry too often, or at the wrong targets, they can say that this shows a culpable failing of moral perception, and not simply a lack of will power.

Those who subscribe to the evaluative view are also much more likely to view moral education as a matter of public concern. The inculcation of correct values plays no necessary role in the mechanistic program, which purports to focus only on behavior. But if the evaluative view is correct, then any program of emotional education that disregards moral

---

121. Hence the more general difference between the Aristotelian and the Kantian views of moral virtue. For Kant virtue is always a matter of strength or mastery, reason dominating basically untrainable impulses. See Kant, *supra* note 20, at 394 ("All strength is known only by the obstacles it can overcome; and in the case of virtue the obstacles are the natural inclinations . . ."). For Aristotle, a condition of struggle against inappropriate inclination gives evidence of moral immaturity; the true state of virtue is one in which the entire personality, including its emotional elements, has internalized an appropriate view of what is right. See Aristotle, *Ethics*, *supra* note 13, at 1103b14–1103b26, 1114b, at 23–25, 51–53; see also M.F. Burnyeat, *Aristotle on Learning to Be Good*, in *Essays on Aristotle's Ethics*, *supra* note 56, at 69, 70–73 (describing, in general terms, the importance of moral education to virtuous actions).

belief is destined to fail. People might succeed for a time in suppressing strong negative impulses—such as racial hatred or homophobia<sup>122</sup> (if these were indeed simply impulses)—but only a person who has learned to discard the ways of seeing African-Americans or gays that give rise to such hate can be relied upon to act stably in accordance with such judgments as circumstances and incentives change.

Indeed, evaluativists have regarded it as an advantage of their conception that it can explain and justify many practices of moral education that do not seem so easily intelligible on the mechanistic model. There is a great deal of room for debate as to how and whether public education and public policy should be used to shape the emotional evaluations that citizens form; many people do think that the only correct focus for law is behavior, and that, so far as the law is concerned, the person who effectively suppresses hatred is on a par with the person who ceases to hate. But even such people would be likely to think of moral education as a shaping of evaluations when considering the role to be played by parents and educators in a child's learning about race. Most of us would not be very satisfied with teachers who ruled racist behavior off-limits but failed to teach the falsity of the beliefs about African-Americans that underlie much racial prejudice and hatred, allowing their students to persist in those appraisals. As parents we tend to treat our children as intelligent creatures whose emotions will develop in accordance with the appraisals they form about the world around them. We consequently attach a great deal of importance to encouraging correct appraisals.

As this discussion should make clear, the evaluative view, while unequivocal in condemning bad character, is ambivalent about how to apportion the blame. The evaluative conception recognizes that a tendency toward certain emotions (for example, the fear of one's own death) may be transmitted, like language-learning ability, as a part of our evolutionary heritage, and that, although it is shaped by social learning, it will be difficult to eradicate completely.<sup>123</sup> Moreover, the recognition that small infants have a life rich in emotion leads to the recognition that early formative experiences may be laid down in the personality in a powerfully influential yet prelinguistic and inchoate form.<sup>124</sup> Such emotions will prove relatively difficult to modify through conscious reflective processes. Finally, social norms shape the emotions of a society's members in such a

---

122. We mean by "homophobia" what psychologists mean by it—namely, an intense disgust or revulsion toward lesbians and gay men. Consistent with the evaluative view, many psychologists treat *cognitive appraisals*—including those arising from past experiences (or more typically, lack of past experiences) with gays or lesbians, anxiety about one's own sexual orientation, and ideological beliefs central to one's conception of self and relationship with others—as essential to this emotion. See Gregory M. Herek, *Beyond "Homophobia": A Social Psychological Perspective on Attitudes Toward Lesbians and Gay Men*, in *Bashers, Baiters & Bigots: Homophobia in American Society* 8–13 (John P. De Cecco ed. 1985).

123. See, e.g., Aristotle, *Ethics*, *supra* note 13, at 1115a5–1116a10, at 53–55.

124. See Lazarus, *supra* note 10, at 180–83.

way that they become deeply habitual;<sup>125</sup> an emotion such as racial fear and hatred, taught in this way, may prove extremely difficult to eradicate, in oneself or in another. Thus the evaluative conception retains a role for suppression when re-education proves too difficult.

But such realistic admissions do not force the holder of this conception to deny that the active contribution of the individual is usually sufficient to make the ethical assessment of emotion appropriate. People are not just the passive pawns either of their parents or of the societies into which they are born; they are capable of critical assessment and reflection, and thus remain obliged to be good even when those around them are not.<sup>126</sup> On the other hand, certain individuals may suffer extremely unfortunate early experiences that severely restrict their power to shape their own characters. In such cases we may not wish to deny responsibility altogether; but we might be inclined to mitigate the punishment we mete out.<sup>127</sup> We shall return to this point in Part IV.

## II. TWO CONCEPTIONS OF EMOTION IN SUBSTANTIVE CRIMINAL LAW

In this Part, we use the mechanistic and evaluative conceptions of emotion to explain various doctrines of substantive criminal law. Although we intend our examination to be primarily descriptive at this stage, our analysis does have a critical objective. In our view, prevailing explanations of these doctrines are inadequate precisely because they overlook the influence of the evaluative conception. Accordingly, we will be arguing in this Part that an account of the law informed by the evaluative conception is superior to any informed solely by the mechanistic view. We begin with an overview of how the two conceptions of emotion influence conventional accounts of moral and legal accountability, and then turn to specific doctrines.

### A. *Emotions and Accountability*

Substantive criminal law is vitally concerned with blame. When is it just to hold a person accountable for her acts? Criminal law theorists

---

125. See Seneca, *On Anger*, *supra* note 15, at 2.20.2–22.2, at 59–61; see also Miriam T. Griffin, *Seneca: A Philosopher in Politics* 125–27 (1976) (arguing that to understand why Seneca stresses *clementia*, one must look to the Rome of his day).

126. It is difficult to say what the contribution of the individual must be in order for assessment to be appropriate, and we shall not embark on that metaphysical question here. Susan Wolf, for example, argues that the criterion of assessability should simply be the ability of the individual to listen to reason, *at the end of the process of development*. This means that two individuals could be equally shaped by their upbringing, and of these one would end up responsible for his character and the other not, according to the end result. See Susan Wolf, *Freedom Within Reason* 94–116 (1990). We neither accept nor reject that idea here; we mention it in order to show the variety of opinions that may be held on the criterion of assessability by people who agree that we may (in most cases) be held responsible for our characters.

127. See Dan M. Kahan & Martha C. Nussbaum, *Emotions Weigh on the Scales of Justice*, *L.A. Times*, July 25, 1995, at B9.

typically give two answers to this question, both of which have strong links with the mechanistic conception of emotion.

*Voluntarism* asserts that individuals are justly held accountable only for willed violations of legal duty. In the words of H.L.A. Hart, the criminal law comprises a "choosing system"; doctrines are and should be structured to make punishment depend on the voluntary decision of an individual to do what the law forbids.<sup>128</sup> The voluntarist approach figures prominently in academic accounts of excuses, but also informs explanations of various other doctrines, including *mens rea*, gradations in homicide, causation, and complicity.<sup>129</sup>

Although voluntarism is not committed analytically to any theory of emotion, it has strong links, both historical and conceptual, with the mechanistic conception. Historically, voluntarism can be traced to Christian (especially Augustinian) doctrines that contrast the will with the forces inherent in our fallen "nature," and to Kant's related view that the will needs persistently to oppose deep-rooted bestial forces or impulses in our nature that can never be internally enlightened or educated. Conceptually, criminal law voluntarists tend to view strong emotions as diminishing an offender's culpability on the ground that they detract from "the accused's capacity for self-control" or constrain her opportunity to exercise it.<sup>130</sup> A person whose "psychological control mechanisms" are overwhelmed by fear or rage cannot justly "be held accountable" for criminal acts.<sup>131</sup> Consistent with the mechanistic conception, emotions enter into such an account only as forces that either do or do not limit an offender's choices; the strength of a person's emotions is thus of far more interest than any valuations internal to them.

*Consequentialism* holds that an individual can be justly punished whenever her behavior frustrates preferred states of affairs.<sup>132</sup> What state of affairs should be maximized is a question that has traditionally divided consequentialists. Following Bentham, however, most criminal law consequentialists rely on a relatively simple wealth-maximizing theory: doctrines should be structured to avoid the greatest amount of harm at the

---

128. See H.L.A. Hart, *Punishment and Responsibility* 46-49 (1968).

129. See, e.g., Hart, *supra* note 128, ch. 2; T.M. Scanlon, *The Significance of Choice* (1986); Stanley I. Benn, *Punishment*, in *Punishment and Rehabilitation* 18, 24-25 (Jeffrie Murphy ed. 1973); Sanford H. Kadish, *Complicity, Cause and Blame: A Study in the Interpretation of Doctrine*, 73 Cal. L. Rev. 323 (1985); Sanford H. Kadish, *Excusing Crime*, 75 Cal. L. Rev. 257 (1987) [hereinafter Kadish, *Excusing Crime*]; Michael Moore, *Choice, Character, and Excuse*, 7 Soc. Phil. & Pol'y 29 (1990) [hereinafter Moore, *Choice*].

130. Hart, *supra* note 128, at 153.

131. 2 Paul H. Robinson, *Criminal Law Defenses* § 177(b)(1) (1984).

132. In this strong form, consequentialism stands in strict opposition to voluntarism. It is, however, possible to combine the two in various ways. See, e.g., Hart, *supra* note 128, ch. 1 (depicting voluntarism as limiting principle on consequentialist theory of punishment); Paul H. Robinson, *Hybrid Principles for the Distribution of Criminal Sanctions*, 82 Nw. U. L. Rev. 19, 28-41 (1988) [hereinafter Robinson, *Hybrid Principles*] (defending complex formula for reconciling competing claims of individual desert and deterrence).

least cost.<sup>133</sup> Other, richer accounts are possible, however;<sup>134</sup> indeed, we ultimately will identify and defend a form of consequentialism that credits the intrinsic worth of emotional valuations. To distinguish the conventional wealth-maximizing version from that one, we will refer to the former as "narrow consequentialism."<sup>135</sup>

Unlike voluntarists, narrow consequentialists do not necessarily view strong emotions as detracting from an offender's culpability. For them, it is an act or disposition's contribution to social welfare, not its origin in a voluntary choice, that governs the attribution of criminal liability.<sup>136</sup> Accordingly, emotions that tend, on average, to frustrate desired states of affairs supply grounds to increase the severity of punishment, whereas those that promote desired states of affairs supply grounds for decreasing it.<sup>137</sup>

However nuanced, this account of emotions is still for all intents and purposes mechanistic. Some narrow consequentialists appear to endorse the mechanistic conception expressly. Richard Posner, for example, distinguishes between "reason" and "emotion," denying the latter any role in identifying courses of action that maximize utility.<sup>138</sup> Posner treats emo-

133. See Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. Pol. Econ. 169 (1968); Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, reprinted in *The Utilitarians* 162 (1961); Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 Colum. L. Rev. 1193, 1204-05, 1209-10 (1985) [hereinafter Posner, *An Economic Theory of the Criminal Law*]; Steven Shavell, *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, 85 Colum. L. Rev. 1232, 1243-46 (1985).

134. See generally Amartya Sen, *Rights and Agency*, 11 Phil. & Pub. Aff. 3 (1981) (defending a conception of consequentialism that includes rights and agent-relative values, and not just social welfare, in the preferred state of affairs).

135. More technically, this form of consequentialism may be characterized as "a species of *welfarist consequentialism* [that] . . . requires simply *adding up* individual welfares or utilities to assess the consequences, a property that is sometimes called *sum-ranking*." Amartya Sen & Bernard Williams, *Introduction*, in *Utilitarianism and Beyond* 1, 4 (Amartya Sen & Bernard Williams eds., 1982). In other words, the preferred state of affairs is that which maximizes welfare, considered as a single sum. In the view of the proponents of law and economics, *welfare* is understood as *wealth*; but it is also possible to advance other related conceptions of welfare.

136. See, e.g., R.B. Brandt, *A Motivational Theory of Excuses in Criminal Law*, in *Nomos XXVII: Criminal Justice* 165, 170-74, 180 (J. Roland Pennock & John W. Chapman eds., 1985).

137. See generally Bentham, *supra* note 133, at 126-27, 144-45, 170 (arguing that emotional dispositions should be appraised according to their tendency to promote or frustrate utility and that punishment should be adjusted to match such propensities); Brandt, *supra* note 136, at 165, 171-76, 187-94 (defending a conception of excuse that connects severity of punishment to the expected utility of an offender's persistent emotional dispositions); Jerome Michael & Herbert Wechsler, *A Rationale of the Law of Homicide II*, 37 Colum. L. Rev. 1261, 1269-70 (1937) (arguing that severity of penalties should be correlated to offender's propensity to engage in dangerous behavior). This approach to assessment of emotions is sometimes referred to as "motive utilitarianism." See, e.g., Robert M. Adams, *Motive Utilitarianism*, 73 J. Phil. 467, 470 (1976).

138. See Richard A. Posner, *The Economics of Justice* 1-2 (1981).

tions simply as powerful forces that push the personality this way or that;<sup>139</sup> criminals who act on such motives, he suggests, can be regarded as "unreasonably dangerous machines."<sup>140</sup>

Moreover, even when they are silent about the nature of emotions, narrow consequentialists employ a distinctively anti-evaluative mode of assessment. Evaluative theorists treat the truth or falsity of the appraisal embodied in an emotion as essential to the emotion's moral status.<sup>141</sup> The narrow consequentialist, in contrast, is only incidentally concerned if at all with this cognitive ingredient of emotion; what matters to her is only whether a particular emotion inclines or disinclines a person to produce the preferred state of affairs.<sup>142</sup>

We will argue in this part that both voluntarism and narrow consequentialism fail, as a descriptive matter, precisely because they are mechanistic. Quite often, criminal law doctrines are structured to assess not the effect of emotion on volition, or the contribution of emotional dispositions to desired states of affairs, but rather the moral quality of the values that a person's emotions express. When this is so, it is possible to make sense of the law only by imputing to it a theory of moral accountability consistent with the evaluative conception of emotion.<sup>143</sup> Aristotle's position on character—that it is appropriate to expect a person to value the right things, in the right ways, at the right times—is one such theory.<sup>144</sup> We will show that this understanding does in fact play a major role in substantive criminal law.

The evaluative position, as we develop it, is much less ambitious than is either the voluntarist or the consequentialist accounts. The evaluative view does not assert that appraising the quality of offenders' emotions is the only object of criminal law or that individuals must invariably be punished or condemned whenever they experience morally inappropriate

---

139. See Richard A. Posner, *Overcoming Law* 186–88 (1995); Posner, *An Economic Theory of the Criminal Law*, *supra* note 133, at 1223 (concluding that impulsive killers may warrant more punishment because impulse intensifies desire to commit crime). This is compatible, though perhaps not easily so, with the position that the *choices* of individuals in matters of sex and emotion are rational in the economic sense. See generally Richard A. Posner, *Sex and Reason* (1992).

140. Richard A. Posner, *The Problems of Jurisprudence* 168–69 (1990).

141. See *supra* text accompanying notes 54–61.

142. See, e.g., Bentham, *supra* note 133, at 100, 126–27, 170 (arguing that emotional dispositions have no intrinsic value but are merely descriptions of settled propensities to behave in particular ways, and that punishment should be adjusted to track the strength and expected utility of such propensities); Posner, *An Economic Theory of the Criminal Law*, *supra* note 133, at 1223 (concluding that impulsive killers may warrant more punishment because impulse intensifies desire to commit crime).

143. For an illuminating analysis that reaches a conclusion similar to ours, see Samuel H. Pillsbury, *Emotional Justice: Moralizing the Passions of Criminal Punishment*, 74 *Cornell L. Rev.* 655 (1989) [hereinafter Pillsbury, *Emotional Justice*]. Pillsbury defends a cognitive conception of emotion, which he argues should inform the law's assessment of individual culpability, particularly in capital sentencing. See *id.* at 674–77, 698–710.

144. See *supra* text accompanying notes 56–57.



emotions. It asserts only that the quality of the offenders' emotion is *one* consideration that matters in substantive criminal law. This claim is admittedly weak. But it is strong enough to pose a substantial challenge to the dominant accounts and is essential, we believe, to making descriptive sense of the law.

We will develop these arguments by considering a range of criminal law doctrines. We start with voluntary manslaughter, which we believe exposes the logic of the evaluative position most clearly. We then examine the doctrines of premeditated murder, self-defense, duress, voluntary act, and insanity.

### B. *Voluntary Manslaughter*

Homicide law typically grades a certain class of emotional killings as voluntary manslaughter rather than murder. The traditional common law formulation—now codified in a majority of American jurisdictions—defines voluntary manslaughter as a killing that is committed in the “heat of passion” produced by an “adequate provocation,” and that occurs without sufficient “cooling time.”<sup>145</sup> A minority of American jurisdictions follow the Model Penal Code, which treats an intentional homicide as manslaughter when “committed under the influence of extreme mental or emotional disturbance.”<sup>146</sup> We will consider these two formulations in turn, examining the extent to which they rest on the mechanistic and evaluative conceptions of emotion.

1. *The Common Law Formulation.* — Most commentators seek to explain the common law formulation of voluntary manslaughter in one of two ways, both of which are mechanistic.<sup>147</sup> The most popular account has voluntarist underpinnings: voluntary manslaughter, under this view, embodies the position that a person who kills in anger or rage has limited culpability because “his choice capabilities [have been] partially undermined.”<sup>148</sup> Consistent with the mechanistic account, emotion mitigates

---

145. See Wayne R. LaFare & Austin W. Scott, Jr., *Criminal Law* § 7.10 (2d ed. 1986).

146. Model Penal Code § 210.3(1)(b) (Official Draft and Revised Comments 1980). See generally Richard Singer, *The Resurgence of Mens Rea: I—Provocation, Emotional Disturbance, and the Model Penal Code*, 27 B.C. L. Rev. 243, 291–304 (1986) (briefly reviewing the history of the Model Penal Code's extreme emotional disturbance standard).

147. There are important exceptions, however. For nonmechanistic accounts, see Jeremy Horder, *Provocation and Responsibility* (1992); Meir Dan-Cohen, *Responsibility and the Boundaries of the Self*, 105 Harv. L. Rev. 959, 993–97 (1992); Pillsbury, *Emotional Justice*, *supra* note 143, at 678–79; and Andrew von Hirsch & Nils Jareborg, *Provocation and Culpability*, in *Responsibility, Character, and the Emotions* 248–51 (Ferdinand Schoeman ed., 1987).

148. Joshua Dressler, *Rethinking Heat of Passion: A Defense in Search of a Rationale*, 73 J. Crim. L. & Criminology 421, 467 (1982) [hereinafter Dressler, *Heat of Passion*]; see also LaFare & Scott, *supra* note 145, § 7.10(b) (“What is really meant by ‘reasonable provocation’ is provocation which causes a reasonable man to lose his normal self-control . . .”).

not because it expresses a morally appropriate evaluation of the actor's situation, but because it impairs her volition.

The second explanation of voluntary manslaughter is rooted in narrow consequentialism. According to this view, individuals who kill in heat of passion upon adequate provocation are both less deterrable and less dangerous than those who kill without provocation or with only minor provocation. Therefore, investing heavily in the punishment of these persons is wasteful.<sup>149</sup> This account is mechanistic because it ties punishment to what an emotion reveals about an actor's propensity to produce undesirable states of affairs rather than to any evaluation expressed by the emotion.

In this section, we show that these accounts are descriptively inadequate precisely because they are mechanistic. An account grounded in the evaluative conception of emotion is superior, in particular, in explaining why the common law requires provocation and what kinds of provocations the common law considers adequate. The evaluative view explains other elements of the doctrine—including the mitigation of punishment for provoked killers, and the requirements of "heat of passion" and "cooling time"—at least as well as the mechanistic accounts. It also supplies a more convincing account of the doctrine than do explanations that depict voluntary manslaughter as either a "justification" or an "excuse."

a. *Adequate Provocation.* — Traces of the mechanistic conception of emotion can definitely be found in the common law authorities. It is not uncommon, for example, for courts to describe the passion presupposed by the doctrine as "blind" and "unreasoning" or as a force that "dominates volition."<sup>150</sup> It might seem quite natural, then, to suppose that the common law approach mitigates the punishment of the impassioned killer precisely because her volition is impaired.

But this voluntarist account fails to make sense of the most basic requirement of the common law formulation: that the defendant's passion arise from a provocation by the victim. It is implausible to say that only provocations can overwhelm an individual's volition; and it is even less plausible to say that an individual whose volition has been overcome will confine her violent outbursts to particular persons. Yet the common law formulation offers no mitigation to the enraged defendant who kills without adequate provocation or who kills someone besides her provoker.<sup>151</sup>

The evaluative conception, in contrast, has no trouble explaining these aspects of the doctrine. There must be provocation because it is only in response to significant slights that anger or rage is morally appro-

---

149. See Brandt, *supra* note 136, at 183–84; Michael & Wechsler, *supra* note 137, at 1280–82.

150. See, e.g., *Disney v. State*, 73 So. 598, 601 (Fla. 1916).

151. See *Cavanaugh v. Commonwealth*, 190 S.W. 123, 126–27 (Ky. 1916); R.S. O'Reagan, *Indirect Provocation and Misdirected Retaliation*, 1968 *Crim. L. Rev.* 319, 323.

priate.<sup>152</sup> And mitigation is warranted only when the angry person attacks her provoker because it is *that* person's conduct that forms the proper object of the evaluation embodied in anger.<sup>153</sup> These features of the doctrine prevent mitigation in circumstances in which the actor's passion, however intense, reflects inappropriate valuations.

Indeed, far from treating impairment of volition as sufficient to mitigate, the early common law authorities consistently rejected that claim precisely because it would indulge "evil passions."<sup>154</sup> A person of "a cruel, vindictive, and aggressive disposition, will seize upon the slightest provocation to satisfy his uncontrolled passions by forming a design to kill."<sup>155</sup> Accordingly, a rule that defines the adequacy of provocation only in terms of the intensity of the defendant's passion would "make evil the palliative of crime, and vice . . . an excuse for its own fruits."<sup>156</sup> By insisting on adequate provocation, the common law formulation confines mitigation to killings that "proceed [not] from a bad or corrupt heart, [but] rather from the infirmity of passion to which even *good men* are subject."<sup>157</sup>

Consistent with the evaluative conception of emotion, this account assumes that the quality of a person's character, not simply the quality of her volition, is the touchstone of moral assessment.<sup>158</sup> A person is the moral author of even those acts that spring from unwilled passions, for she remains accountable for experiencing good rather than bad emotions.<sup>159</sup> A rule that focused only on impairment of volition would thus

152. See, e.g., *Regina v. Welsh*, 11 Cox Crim. Cas. 336, 337 (1869) (rejecting claim that "influence of ungovernable passion" would support finding of involuntary manslaughter where "provocation was . . . slight").

153. See von Hirsch & Jareborg, *supra* note 147, at 254 n.40.

154. *Maher v. People*, 10 Mich. 212, 220 (1862).

155. *Rivers v. State*, 78 So. 343, 345 (Fla. 1918).

156. *Small v. Commonwealth*, 91 Pa. 304, 308 (1879); see also *People v. Logan*, 164 P. 1121, 1122 (Cal. 1917):

[N]o defendant may set up his own standard of conduct and justify or excuse himself because in fact his passions were aroused, unless further the jury believe that the facts and circumstances were sufficient to arouse the passions of the ordinarily reasonable man. Thus no man of extremely violent passion could so justify or excuse himself if the exciting cause be not adequate, nor could an excessively cowardly man justify himself unless the circumstances were such as to arouse the fears of the ordinarily courageous man.

157. *State v. Cook*, 3 Ohio Dec. Reprint 142, 144 (1859) (emphasis added).

158. See *Small*, 91 Pa. at 308 ("Man is largely the creature of education; his character depends principally, if not wholly, upon his own will, and for that character he is legally responsible.").

159. See *id.* ("Suppose then we admit testimony that the defendant is quick-tempered, violent and revengeful; what then? Are these an excuse for, or do they even mitigate crime? Certainly not, for they result from a want of self-discipline; a neglect of self-culture that is inexcusable."); *Keenan v. Commonwealth*, 44 Pa. 55, 59 (1862) ("[M]en who degrade themselves below the ordinary level of social morality, by bad conduct or habits, do not thereby relieve themselves from having their acts and duties judged by the ordinary rules of social action. They cannot set up their own vices as a reason for being set into a special class that is to be judged more favourably than other persons.").

risk getting things entirely backwards; it would show more solicitude toward the "proud, or captious, or selfish or habitually ill-natured man"—"who by any sort of indulgence, fault, or vice, renders himself very easily excitable, or very subject to temptation"—than it would toward the "moderate, well-tempered, and orderly citizen."<sup>160</sup>

The common law formulation requires not only that there be a provocation but that the provocation be legally *adequate*. The early common law authorities defined as a matter of law the circumstances in which an intentional killing could be mitigated to voluntary manslaughter.<sup>161</sup> The distinctions between provocations that were adequate and those that were not were often quite fine. For example, a blow to the face was adequate, a boxing of the ears not;<sup>162</sup> the infidelity of a man's wife was adequate, the infidelity of a man's fiancée or girlfriend not.<sup>163</sup>

These categories are difficult to explain under the voluntarist view. If the morally salient feature of a provoked individual's circumstances is loss of control, why aren't all volition-impairing provocations deemed to be "adequate"? It is sometimes suggested that the common law categories were substitutes for factual inquiries into volitional impairment, which is very difficult to measure in actual cases.<sup>164</sup> But if they were merely generalizations about the kinds of offenses that typically destroy volition, the common law categories seem manifestly underinclusive. It seems implausible, for example, to think that nineteenth-century decisionmakers would have viewed rage as a surprising or abnormal reaction to the infidelity of a man's fiancée.

In fact, these authorities clearly conceived of the categories in evaluative terms. The categories embody judgments about what kinds of goods are appropriately valued and by whom. The common law authorities, for example, viewed adultery as "the gravest possible offence which a wife can commit against her husband"<sup>165</sup> and "the highest invasion of [his] prop-

---

160. *Keenan*, 44 Pa. at 58; see also *Maher v. People*, 10 Mich. 212, 220 (1862) (if impairment of volition were sufficient to show adequacy of provocation, "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").

161. See LaFave & Scott, *supra* note 145, § 7.10(b).

162. Compare *Sir Michael Foster*, Crown Cases 292 (1809) (ear-boxing) with *Stewart v. State*, 78 Ala. 436, 440 (1885) (blow to face).

163. Compare *Regina v. Mawgridge*, 84 Eng. Rep. 1107, 1115 (1707) (discovery of adultery constitutes legally adequate provocation) with *Rex v. Palmer*, 2 K.B. 29, 30-31 (1913) (discovery of fiancée's infidelity is not legally adequate provocation).

164. See, e.g., Hart, *supra* note 128, at 33 ("Further difficulties of proof may cause a legal system to limit its inquiry into the agent's 'subjective condition' by asking . . . whether a 'reasonable man' would have been deprived (say, by provocation) of self-control . . ."); Herbert Wechsler & Jerome Michael, *A Rationale of the Law of Homicide*: I, 37 Colum. L. Rev. 701, 718 (1937) (speculating that common law categories of adequate provocation were "apparently [based] on the ground that they alone resulted in emotional pressure under which ordinarily law abiding men might be expected to collapse").

165. *Rex v. Greening*, 3 K.B. 846, 849 (1913).

erty" by another man.<sup>166</sup> The infidelity of an unmarried woman, however, was "entirely different," for "the man has no such right to control the woman as a husband has to control his wife."<sup>167</sup> The law must thus treat an enraged man who kills his girlfriend's lover differently from an enraged man who kills his wife's paramour (even if both men are to be punished), not because their emotions are different in intensity, but because their emotions reflect valuations of honor and dignity that it would be morally obtuse to equate.<sup>168</sup>

Modern authorities have tended to abandon categorical definitions of adequate provocation,<sup>169</sup> but for reasons that are also perfectly consistent with the evaluative understanding. The contemporary approach recognizes, first, that the circumstances in which provocations are adequate cannot be fully specified in advance. The considerations relevant to evaluating a provoked individual's emotions are too diverse and too particular to be reduced to any set of categories or rules:

[The adequacy of provocation] must vary with, and depend upon the almost infinite variety of facts presented by the various cases as they arise. The law can not with justice assume, by the light of past decisions, to catalogue all the various facts and combinations of facts which shall be held to constitute reasonable or adequate provocation. . . . Provocations will be given without reference to any previous model, and the passions they excite will not consult the precedents.<sup>170</sup>

Second, the contemporary approach assumes that juries are generally better at making fact-specific appraisals of defendants' emotions than

166. *Mawgridge*, 84 Eng. Rep. at 1115.

167. *Greening*, 3 K.B. at 849.

168. See *id.* The "ear boxing"- "blow to the face" distinction also appears to have been rooted in evaluative notions, at least at its inception. Foster describes *Stedman's Case*, from which the rule derives, this way:

There being an affray in the street, one Stedman, a footsoldier, ran hastily towards the combatants. A woman seeing him run in that manner cried out, 'You will not murder the man will you?' Stedman replied, 'What is that to you, you bitch?' The woman thereupon gave him a box on the ear, and Stedman struck her on the breast with the pommel of his sword. The woman then fled, and Stedman pursuing her stabbed her in the back. Holt was at first of opinion, that this was murder, a single box on the ear from a woman not being a sufficient provocation to kill in this manner, after he had given her a blow in return for the box on the ear . . .

Foster, *supra* note 162, at 292. This statement of the rule highlights the defective conception of honor embodied in the rage of a man who responds to the taunts of a defenseless woman by stabbing her in the back.

169. See generally LaFave & Scott, *supra* note 145, § 7.10(b) (noting modern trend away from pigeon-holing provocative conduct). But cf. Sanford H. Kadish & Stephen J. Schulhofer, *Criminal Law and Its Processes: Cases and Materials* 413 (6th ed. 1995) (suggesting that "[m]ost modern courts are only slightly more flexible" than were common law authorities in recognizing provocations as adequate).

170. *Maier v. People*, 10 Mich. 212, 221-22 (1862) (citations omitted); accord *State v. Gounagias*, 153 P. 9, 12 (Wash. 1915); *Campbell v. Commonwealth*, 11 S.W. 290, 292 (Ky. 1889).

are judges. "[C]oming from the various classes and occupations of society, and conversant with the practical affairs of life," jurors can more legitimately declare the adequacy of a provocation than can a judge, "whose habits and course of life give him much less experience of the workings of passion in the actual conflicts of life."<sup>171</sup> Thus, whereas the early authorities constructed rigid legal categories to reflect what they perceived as objectively grounded evaluations, contemporary authorities make the adequacy of provocation an issue of fact for the jury so that the law may assess emotions against the background of community mores.

Nevertheless, while many courts no longer purport to specify all the provocations that are adequate as a matter of law, they still occasionally identify particular ones that are not. Some (but not all) courts, for example, have refused to permit defendants to present voluntary manslaughter theories in cases in which their victims have made homosexual advances toward the defendant or engaged in similar behavior.<sup>172</sup> These decisions, too, are best understood in evaluative rather than mechanistic terms. They do not assume that the asserted provocations were insufficient to destroy the defendants' volition; indeed, many of these cases have excluded expert psychiatric testimony designed to show exactly that.<sup>173</sup> Rather, they deem the provocations insufficient because they conclude that the law should criticize rather than endorse the evaluation of the victim's identity implicit in the defendant's rage.<sup>174</sup>

This evaluative account is also descriptively superior, we believe, to the mechanistic view implicit in the narrow consequentialist account of the doctrine. The consequentialist explanation posits that individuals who kill only upon a significant provocation are not sufficiently dangerous to warrant the severe punishment ordinarily imposed on other intentional killers.<sup>175</sup> However, the common law paradigms of "adequate provocation"—adultery and humiliating but non-life-threatening blows to the face—occur frequently without leading to deadly retaliation. It thus seems strange to suggest that the law mitigates because persons who intentionally kill in response to such transgressions do not seem as life-threatening as those who intentionally kill for other reasons.

Perhaps a person who kills once under such conditions is unlikely to kill again. But how can a consequentialist be sure that the savings from relaxing one offender's punishment on this ground will not be offset by the deterrence-undermining effects of such a disposition on the behavior of others who find themselves in such situations for the first time? In-

---

171. *Maier*, 10 Mich. at 221.

172. See *Commonwealth v. Carr*, 580 A.2d 1362, 1364–65 (Pa. Super. Ct. 1990); Robert B. Mison, Comment, Homophobia in Manslaughter: The Homosexual Advance as Insufficient Provocation, 80 Cal. L. Rev. 133 (1992); Developments in the Law—Sexual Orientation and the Law, 102 Harv. L. Rev. 1508, 1547 (1989).

173. See, e.g., *Carr*, 580 A.2d at 1363–65.

174. See *id.* at 1364–65.

175. See *supra* text accompanying note 149.

deed, if the goal of the law were really to maximize savings in life, it would likely punish killings in these settings more severely, not less, precisely because the factual circumstances are common and are likely to spark particularly intense desires for retaliation.<sup>176</sup>

The narrow consequentialist account does little better in explaining the prevailing contemporary approach, which abandons rigid legal categories. Because this approach leaves juries free to apply their own evaluative judgments, the consequentialist theory holds up only if we imagine that juries (consciously or unconsciously) always rank the reprehensibility of defendants' passions according to the tendency of particular emotions to promote killings. This claim also seems implausible. No doubt, juries convict some impulsive killers of murder rather than voluntary manslaughter precisely because these killers' anti-social emotions make them seem exceedingly dangerous. But juries also routinely convict only of manslaughter other impassioned defendants who kill in response to common and recurring impositions—from adultery to domestic violence.<sup>177</sup> Again, one would expect the latter class of defendants also to be convicted of murder if voluntary manslaughter verdicts tracked the propensity of different types of emotions to promote killings.<sup>178</sup>

---

176. See C.L. Ten, *Crime, Guilt, and Punishment* 147–48 (1987); Dressler, *Heat of Passion*, *supra* note 148, at 434 n.136; cf. Posner, *An Economic Theory of the Criminal Law*, *supra* note 133, at 1223 (noting that deterrence theory may justify greater punishment to compensate for strength of impulsive killers' desire to kill).

177. See, e.g., Harry Kalven, Jr. & Hans Zeisel, *The American Jury* ch. 16 (1966). According to one recent empirical study, members of the public are readily disposed to downgrade the seriousness of killings provoked by offensive but nondeadly behavior. See Paul H. Robinson & John M. Darley, *Justice, Liability & Blame: Community Views and the Criminal Law* 56–57 (1995).

178. A stronger—and even less plausible—consequentialist claim would be that the law punishes such persons less severely because their propensities deter wrongdoing by others. See, e.g., Alon Harel, *Efficiency and Fairness in Criminal Law: The Case for a Criminal Law Principle of Comparative Fault*, 82 Cal. L. Rev. 1181, 1215–16 (1994). It might be individually rational for a person to cultivate a disposition to react violently to transgressions; such a person is less likely to be crossed than one who has a reputation for submitting meekly to indignities. See Robert H. Frank, *Passions Within Reason* 5 (1988); David D. Friedman, *Price Theory* 290 (2d ed. 1990). But it almost certainly would not be collectively rational for society to propagate a vengeful disposition among its members generally, for then violent resolution of disputes would be much too common. See Jon Elster, *The Cement of Society: A Study of Social Order* 144, 283 (1989); Bentham, *supra* note 133, at 137–38; see also Frank, *supra*, at 39 (noting that cultural norms generally discourage vengefulness); cf. Morton Deutsch & Robert M. Krauss, *The Effect of Threat upon Interpersonal Bargaining*, 61 J. Abnormal & Soc. Psychol. 181 (1960) (reporting results of experimental study suggesting that it is individually rational but collectively irrational to resort to threats in negotiations). Indeed, empirical studies suggest that the incidence of violent deaths is extremely high in societies with strong retribution norms. See Elster, *supra*, at 284; Napoleon A. Chagnon, *Life Histories, Blood Revenge, and Warfare*, 239 Sci. 985 (1988); see also Christopher Boehm, *Blood Revenge* 175–80 (1984) (discussing effect of feuding on population sizes). So it seems very unlikely that the contours of voluntary manslaughter doctrine can be imputed to the economic efficiency of rage.

Of course, the consequentialist need not view *lives* per se as what voluntary manslaughter is maximizing. The law can be viewed as trying to maximize particular states of affairs that entail the full protection of only certain lives.<sup>179</sup> From the common law provocation categories, for example, it can be inferred that the law does not attach as much value to the life of the paramour as it does to the life of the average person, and for that reason it invests less in punishing emotions that promote killings of the former. Jury verdicts under the contemporary approach can be viewed as revealing similar assessments of the relative values of different persons' lives. Once the law is fully decoded to identify the states of affairs that are truly valued, the consequentialist might argue, it will be seen that voluntary manslaughter doctrine does consistently assess emotions according to whether they promote or frustrate valued states of affairs.<sup>180</sup>

But this move rescues narrow consequentialism at too high a cost. The value this approach assigns to states of affairs is completely derivative from the assessments that legal decisionmakers (whether courts or juries) make of the evaluations embodied in defendants' emotions: it determines that the law attaches low value to the life of the paramour because decisionmakers are prepared to endorse the husband's anger as reasonable; it doesn't say that the law endorses the husband's anger because his emotion is consistent with an independently specified valuation of the paramour's life. Thus, whereas consequentialists purport to explain criminal law by identifying the states of affairs that are normative for emotions, they can in fact identify valued states of affairs only by crediting the measures of value internal to the emotions themselves. The result is a much richer form of consequentialism that values or disvalues emotional appraisals for their own sake and not for their propensity to maximize social wealth or other independently specified states of affairs. Such an approach presupposes an evaluative conception of emotion.

b. *Mitigation vs. Exculpation.* — Voluntary manslaughter only mitigates punishment and does not fully exculpate the impassioned killer. The mechanistic accounts supply explanations for this feature of the doctrine: under the voluntarist position, mitigation is appropriate because the doctrine contemplates a quantum of passion that only "partially undermine[s]" a person's "choice-capabilities";<sup>181</sup> under the narrow consequentialist position, provoked killers are still worth punishing, but they

---

179. Cf. Michael & Wechsler, *supra* note 137, at 1271-72 (noting that value of deterring homicide increases as value of life threatened by homicide increases).

180. See Brandt, *supra* note 136, at 174 (suggesting that the law can be viewed as tailoring severity of punishment to "defects of motivation," which can be derived from "those stated or implied by the prohibitions (in statutes or precedents) of a given legal system").

181. Dressler, *Heat of Passion*, *supra* note 148, at 467; see also Hart, *supra* note 128, at 153 (where offender is provoked the law "punish[es] less, on the footing that, though the accused's capacity for self-control was not absent its exercise was a matter of abnormal difficulty"); LaFave & Scott, *supra* note 145, § 7.10(b) (equating "reasonable provocation" with "provocation which causes a reasonable man to lose self-control").



are not worth as much punishment as the (supposedly) more dangerous class of unprovoked killers.<sup>182</sup>

Why the doctrine mitigates and does not fully exculpate might seem more difficult to explain under the evaluative view. If the defendant's anger or rage embodies a morally appropriate evaluation of the victim's provoking conduct, why does the law not fully immunize the defendant from condemnation? Alternatively, if killing the victim was morally wrong, how can it be said that the defendant's anger embodied a morally appropriate evaluation? If the killing was wrong, shouldn't the law, on the evaluative view, fully condemn the homicide as murder?

We think that there is a ready solution to this puzzle. The evaluative view does not assume that any particular act expresses only one emotional evaluation, which is either "appropriate" or "inappropriate." Acts may be attended by multiple evaluations varying in their moral quality. Thus, for the law to be evaluative in nature, it is enough for it to condemn acts that reflect at least some appropriate emotional motivations less severely than acts that reflect only inappropriate emotional motivations.

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.<sup>183</sup> 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.

Now compare this woman to a man who, like Carr, impulsively kills out of homophobic hatred.<sup>184</sup> He is differently situated. Whereas condemnation of the mother is likely to be qualified by our recognition of the appropriateness of her anger, the proper moral attitude toward the

---

182. See *supra* note 149 and accompanying text.

183. See, e.g., *People v. Shields*, 575 N.E.2d 538, 546 (Ill. 1991) (victim's admission that he had raped defendant's daughter and statement that he would do so again, in combination with victim's physical assault of defendant, constitutes sufficient provocation).

184. See *Commonwealth v. Carr*, 580 A.2d 1362, 1364 (Pa. Super. Ct. 1990) (concluding as a matter of law that aversion to homosexuality is insufficient provocation).

homophobe (we submit)<sup>185</sup> is one of unqualified outrage informed in large measure by revulsion toward the false appraisal of the victim's worth expressed in the offender's hatred.<sup>186</sup> If the law seeks to take account of the evaluations that inform emotions, then it should not treat these two offenders the same; it ought to punish both, but it ought to punish the homophobe more severely, precisely because the falsity of the judgments expressed in his hatred merits stronger condemnation than does the very justifiable anger of the parent.

This is exactly how the common law authorities understand the mitigating upshot of voluntary manslaughter. The common law authorities reduced the punishment of the man who killed his wife's paramour not because *killing* was the morally appropriate thing to do, but because the husband's *appropriate anger* distinguished him, morally, from others who kill without appropriate passions. Similarly, the common law refused to mitigate the punishment of a man who killed the lover of his girlfriend or fiancée because that man's rage expressed a valuation of honor and dignity that could not properly be equated with that of the cuckold.<sup>187</sup>

The same form of reasoning is reflected more generally in the common law's distinction between murder and manslaughter. Murder requires proof of "malice aforethought," which conveys that the "act [was] prompted by, or . . . sprung from, a wicked, depraved or malignant mind—a mind which, even in its habitual condition, and when excited by no provocation . . . is cruel, wanton or malignant, reckless of human life, or regardless of social duty."<sup>188</sup>

But if the act of killing, though intentional, be committed under the influence of passion or in heat of blood, produced by an adequate or reasonable provocation . . . rather than of any wickedness of heart or cruelty or recklessness of disposition; then the law, out of indulgence to the frailty of human nature . . . very properly regards the offense as of a less heinous character than murder, and gives it the designation of manslaughter.<sup>189</sup>

In both classes, killing is wrong and worthy of punishment; it is simply "less heinous" in the case of manslaughter. And it is less heinous precisely because the quality of a person's emotions affects the moral assessment of her acts. The common law insists on distinguishing "manslaugh-

185. Some might disagree; indeed, some courts have. See generally Mison, *supra* note 172, at 133 (canvassing legal responses). The analysis remains evaluative, however, even when the evaluation is inverted.

186. See generally Richard A. Berk et al., Thinking More Clearly About Hate-Motivated Crimes, in *Hate Crimes: Confronting Violence Against Lesbians and Gay Men* 123, 127–28, 129 (Gregory M. Herek & Kevin T. Berrill eds., 1992) (identifying symbolic and expressive dimensions of homophobic violence); Karl M. Hamner, *Gay Bashing: A Social Identity Analysis of Violence Against Lesbians and Gay Men*, in *Hate Crimes*, *supra*, at 179, 179–83 (concluding that aim of homophobic violence is to improve social status of heterosexuals by lowering that of homosexuals).

187. See *supra* text accompanying notes 165–168.

188. *Maher v. People*, 10 Mich. 212, 218 (1862).

189. *Id.* at 218–19; accord *State v. Gounagias*, 153 P. 9, 12 (Wash. 1915).

ter" from "murder" because it would be morally obtuse to equate a "passion . . . produced by an adequate or reasonable provocation" with the "wickedness of heart or cruelty" associated with "malice aforethought."

c. *Heat of Passion*. — Many commentators cite the "heat of passion" requirement to show that the common law formulation mitigates on account of impaired volition.<sup>190</sup> But this argument incorrectly assumes that a person's emotions can affect the moral assessment of her acts only in the way that the mechanistic theory suggests. Under an evaluative explanation, too, the existence of anger, rage, or fear is properly made a condition of mitigation and for reasons that have nothing to do with any impairment of volition caused by such a passion.

To begin, the evaluative view corresponds to a theory of moral assessment that is concerned not just with what an individual does but with why she does it. A person can perform a desirable act for a variety of reasons; according to the evaluative view, that act has greatest moral worth when it is the product of motives (including emotions) that themselves express an appropriate evaluation of the actor's circumstances.<sup>191</sup> Similarly, an undesirable act can be carried out for a variety of reasons and is more or less worthy of condemnation depending on what the actor's motives for doing that act express.

Because it focuses on motives and not just outcomes, the evaluative view is perfectly able to explain why the traditional doctrine confines mitigation to those who experience "heat of passion." Without that emotion, it would be impossible to understand the defendant's act as expressing an appropriate valuation of the good—whether it is the defendant's honor or the dignity or physical security of the defendant's family members—that is threatened by the victim's wrongful provocation. In particular, without anger or rage, 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."<sup>192</sup>

In addition, the evaluative view is concerned with *relative valuations*. A good person not only values the right things but values them in the right amount in relation to other goods. We might expect a person who is snubbed by a colleague in a professional setting to be angry; anger

---

190. See Joshua Dressler, *Provocation: Partial Justification or Partial Excuse?*, 51 *Mod. L. Rev.* 467, 479–80 (1988) [hereinafter Dressler, *Provocation*]; Singer, *supra* note 146 at 312–14; cf. von Hirsch & Jareborg, *supra* note 147, at 253 (arguing that heat of passion requirement should be dropped if manslaughter is viewed as resting on defendant's appropriate resentment of mistreatment).

191. See Elizabeth Anderson, *Value in Ethics and Economics* 33–34 (1993); Nussbaum, *Therapy*, *supra* note 15, at 392; Michael Stocker, *Plural and Conflicting Values* 114 (1990).

192. *Maher*, 10 Mich. at 220; accord *Gounagias*, 153 P. at 12.

reveals that she properly values her honor and dignity. At the same time, if this person became *more* angry at this slight than at, say, the wrongful infliction of an injury upon her child, we would say that her relative valuations are skewed; the intensity of her anger would then reveal that her love of honor is excessive in relation to her love of her child.

The "heat of passion" requirement accommodates assessments of an offender's relative valuations. The existence of passion demonstrates that the offender values the good (which, of course, must be something that a person of good character would value) sufficiently in relation to other goods. If a man dispassionately killed his wife's paramour—much as he might dispassionately kill an annoying mosquito—we would suspect that his beliefs about what is important are skewed: the absence of anger would show us that he invests too little value in fidelity; his act of killing without anger would show us that he invests too little value in others' lives.<sup>193</sup>

Thus, it is relatively easy to make sense of the "heat of passion" requirement in evaluative terms; indeed, it is somewhat difficult, on close inspection, to make complete sense of this requirement in purely mechanistic ones. The common law authorities insist on "heat of passion" but they do not insist that the defendant's passion be of any particular intensity—much less of volition-distorting intensity.<sup>194</sup> 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*."<sup>195</sup> This conception of "heat of passion" is not impossible to reconcile with the mechanistic account, which suggests (somewhat obscurely) that "partial impairment" of volition is adequate to mitigate "partially."<sup>196</sup> Nevertheless, the common law's de-emphasis of volitional impairment seems much more consistent with the position that certain types of cognitive evaluations are in fact sufficient conditions of emotions.<sup>197</sup>

d. *Cooling Time*. — At first glance, the "cooling time" limitation also appears mechanistic. Courts frequently describe it as the interval in which "blood" can be expected "to cool," "temporary excitement" to abate, and "reason to reassert itself."<sup>198</sup> Such language might be read as suggesting an image of emotion as a shock or jolt that grips a person and compels her to act but that eventually passes on. For this and other reasons, it might be thought that any notion of "cooling time" is alien to the evaluative conception: why would anger that is deemed reasonable at one time be any less so later on? And if the anger is based on a thought

---

193. We are grateful to Dick Craswell for helpful discussion on this point.

194. See A.J. Ashworth, *The Doctrine of Provocation*, 35 Cambridge L.J. 292, 306 (1976).

195. *Maher*, 10 Mich. at 220.

196. See *supra* note 181 and accompanying text.

197. See *supra* text accompanying notes 98–104.

198. *State v. Gounagias*, 153 P. 9, 12 (Wash. 1915); see also *Maher*, 10 Mich. at 222–24.

about the wrong that has occurred, and this thought is correct, why should it go away after a time? Nevertheless, we believe that the evaluative conception of emotion ultimately makes more sense of the "cooling time" limitation than does the mechanistic conception.

The mechanistic view can perhaps explain the rhetoric of "cooling time," but it cannot explain the substance of it. The mechanistic account implies that after a relatively short period of time the "forces" and "drives" of anger no longer operate in the person's psyche; reason necessarily reasserts its control by suppressing a drive or force that is by its very nature not a thinking or reasoning force, in effect ordering it to subside. But this is not how the doctrine is typically conceived of by courts, which frequently find sufficient "cooling time" even when it is conceded that the defendant remained in a state of intense agitation. Consider *People v. Ashland*,<sup>199</sup> in which the court concluded, as a matter of law, that seventeen hours was sufficient "cooling time" for a man's rage to be brought under control following discovery of adultery:

Undoubtedly [the defendant] became very angry and perhaps much beside himself upon receiving this information, and undoubtedly he remained in a high state of anger up to the time that he gratified his resentment of [the paramour's] acts by killing him. And it is very probable that, had he not met the deceased after hearing of his wife's infidelity, he would for many years thereafter and perhaps for the remainder of his life have been aroused to a state of intense passion whenever the acts of [the victim] recurred to him; yet no one would say that he could avenge the wrong committed against him by [the victim] by killing the latter at any time he might happen to meet him after learning of such wrong and after the lapse of plenty of time for his passion to give way to the normal tranquility of his judgment . . . .<sup>200</sup>

It is much easier to make sense of "cooling time" in evaluative terms. We assess the appropriateness of emotions against a complex of social norms. Some of these, as we've explained, specify which goods are valued by a reasonable person, and how much a reasonable person should value such goods in relation to each other. Still other norms, however, specify how long a person should persist in a state of strong or intense anger when someone has interfered with some good essential to his well-being.<sup>201</sup>

Indeed, as *Ashland* highlights, the duration of a person's anger is particularly relevant to the assessment of his relative valuations. When an offense is fresh, we tolerate—indeed, expect—strong anger; it is precisely because a person *ought* to value fidelity intensely that the law is prepared to mitigate the punishment of the cuckold if he kills shortly after learning

---

199. 128 P. 798 (Cal. Ct. App. 1912).

200. Id. at 802.

201. See Averill, *supra* note 65, at 278–79.

of the adultery. But if 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”<sup>202</sup> in a way that reflects an appropriate valuation of all the goods and interests at stake.

This said, we are fully prepared to concede that the evaluative account of “cooling time” is neither perfect nor exclusive. It is not hard to imagine cases in which a defendant's smoldering or brooding anger seems morally appropriate or understandable, but in which a court might nonetheless be constrained as a matter of law, or even a jury as a matter of fact, to find “cooling time.” But this lingering imprecision is probably best attributed to the inherent impossibility of devising legal rules that perfectly capture the relevant moral judgments in this setting. In particular, as time passes after a wrongful provocation, it becomes increasingly difficult to determine whether the defendant's action was a genuine impassioned response to the provocation or a killing carried out for some other reason. The flexible conception of “cooling time” reflected in case law works adequately well for evaluative purposes and otherwise marks a reasonable accommodation between the evaluative view and other legitimate concerns in criminal law, including deterrence.

e. *Justification or Excuse?* — Some of the doctrinal points that we have tried to relate to competing conceptions of emotion figure prominently in academic debates over whether the common law formulation is best characterized as a “justification” or as an “excuse.”<sup>203</sup> Our analysis suggests that the answer to this question is “neither”; the terms of this debate are themselves inadequate because they embody theories of moral assessment that assume a mechanistic rather than an evaluative conception of emotion.

It is widely assumed that “justification” and “excuse” embody normative distinctions that generalize across criminal defenses.<sup>204</sup> Justifications

---

202. *Ashland*, 128 P. at 802.

203. Compare Ashworth, *supra* note 194, at 307 (justification) and Finbarr McAuley, *Anticipating the Past: The Defence of Provocation in Irish Law*, 50 *Mod. L. Rev.* 133, 139–40 (1987) (same) with Dressler, *Provocation*, *supra* note 190, at 472 (excuse) and Singer, *supra* note 146, at 306 (same).

204. See, e.g., J.L. Austin, *A Plea for Excuses*, 57 *Proc. Aristotelian Soc'y* 1 (1956–1957), reprinted in J.L. Austin, *Philosophical Papers* 175, 175–76 (J.O. Urmson & G.J. Warnock eds., 3d ed. 1979); George P. Fletcher, *Rethinking Criminal Law* ch. 10 (1978) [hereinafter Fletcher, *Rethinking Criminal Law*]; Paul H. Robinson, *A Theory of Justification: Societal Harm as a Prerequisite for Criminal Liability*, 23 *U.C.L.A. L. Rev.* 266, 274–75 (1975) [hereinafter Robinson, *Justification*]. For a challenge to the utility of

are said to identify acts that produce morally preferred states of affairs.<sup>205</sup> When complying with the law would result in greater harm than would breaking it, for example, a defendant may assert the justification of "necessity";<sup>206</sup> an individual who kills another to protect his own life may assert the justification of self-defense, because the law prefers the death of the wrongful aggressor to the death of the law-abiding citizen.<sup>207</sup> In addition, justifications are said to be "universal" and "objective"; because they single out acts that produce superior states of affairs, justifications are indifferent to the identity of the actor or her motive for doing the act.<sup>208</sup>

Excuses, in contrast, are said to identify circumstances in which an act is wrongful but the actor blameless.<sup>209</sup> For example, a person whose will is overborne by a threat may be able to assert the excuse of "duress" even if her act imposes a greater harm than is threatened to her.<sup>210</sup> Moreover, excuses are said to be "subjective" and "individualized."<sup>211</sup> They are concerned with how the defendant's particular circumstances affected her capacity or opportunity to obey the law.<sup>212</sup>

The common law formulation of voluntary manslaughter does not fit neatly into either of these categories. Consider again the parent who is convicted of manslaughter after killing a man who has abused her child.<sup>213</sup> By mitigating her punishment, the law presumably does not imply that the man's death, by itself, produces a better state of affairs.<sup>214</sup> Moreover, application of the doctrine to the angry parent does not satisfy

---

the distinction, see R. Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 *Colum. L. Rev.* 1897, 1897 (1984).

205. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 769 ("The modern claim is that all justificatory arguments can be reduced to a balancing of competing interests and a judgment in favor of the superior interest."); Paul H. Robinson, *Criminal Law Defenses: A Systematic Analysis*, 82 *Colum. L. Rev.* 199, 213-14 (1982) [hereinafter Robinson, *Systematic Analysis*].

206. See, e.g., Model Penal Code § 3.02(1)(a) (1985) (defense where "the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense charged"); Robinson, *Systematic Analysis*, *supra* note 205, at 213-14.

207. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 857-64; Robinson, *Systematic Analysis*, *supra* note 205, at 236.

208. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 762; Robinson, *Justification*, *supra* note 204, at 274-75. But cf. Greenawalt, *supra* note 204, at 1915-16 (showing that justifications do sometimes take the defendant's role and motive into account).

209. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 798; Robinson, *Systematic Analysis*, *supra* note 205, at 221-22.

210. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 831.

211. See *id.* at 762. But cf. Greenawalt, *supra* note 204, at 1915-18 (observing that the duress defense has an objective component in most jurisdictions).

212. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 802-07; Moore, Choice, *supra* note 129, at 31-40; Robinson, *Justification*, *supra* note 204, at 275; Robinson, *Systematic Analysis*, *supra* note 205, at 221-22.

213. See *People v. Shields*, 575 N.E.2d 538, 545-46 (Ill. 1991).

214. The man, for example, could not have been executed by the state for his behavior. See *Coker v. Georgia*, 433 U.S. 584, 592-93 (1977).

the objectivity and universality conditions of justification: it was necessary for the parent to show that she killed out of anger and not for some other reason; and certainly no other person could have killed on the parent's behalf and still have been entitled to present a manslaughter theory. Nor does voluntary manslaughter appear to satisfy the conditions of "excuse" in this setting. It would not be necessary for the parent to show that her anger deprived her of the capacity to obey the law; but even if she could, that would not be sufficient unless she could also establish that the victim's acts were adequate to provoke a reasonable person.

If it is assumed that all defenses must satisfy the conditions of either justification or excuse, then voluntary manslaughter doctrine appears unprincipled and worthy of reform.<sup>215</sup> But why make this assumption? If there is a coherent explanation for voluntary manslaughter, then the antagonism between the doctrine and the categories of "justification" and "excuse" would reflect only the descriptive poverty of these concepts.

The evaluative conception of emotion supplies such an explanation. Under the evaluative view, the contours of the common law doctrine make perfect sense because they allow the law to track the quality of the evaluations expressed in intentional killers' emotions. The angry parent is entitled to the mitigating consequences of the doctrine not because her act produced the best state of affairs or because her anger deprived her of control, but rather because her anger was appropriate for someone in her situation. The appropriateness of her emotional motivations, moreover, distinguish her from a person who kills on the basis of less appropriate or fully inappropriate motives.

The reason that "justification" and "excuse" fail to explain the doctrine is that these concepts as currently understood presuppose the theories of moral assessment associated with the mechanistic conception of emotion. When connected to actions that promote preferred states of affairs, "justification" assumes a narrow consequentialist theory of assessment; this understanding of justification makes no allowance for assessments that focus on the quality of the actor's motives (including her emotional motives) abstracted from the consequences of her acts.<sup>216</sup> The prevailing understanding of excuse does focus on the actor's subjective motivation, but only to determine whether her acts are freely chosen; because emotions are unwilling, this theory of assessment furnishes no ground for distinguishing among actors based on the quality of their pas-

---

215. For suggestions to recast the doctrine as "pure" excuse, see Dressler, *Heat of Passion*, *supra* note 148, at 424; O'Reagan, *supra* note 151, at 323-24; Singer, *supra* note 146, at 294. For a proposal to recast the doctrine as a "partial justification," see McAuley, *supra* note 203, at 139-42, 156-57.

216. Of course, a richer form of consequentialism—one, for example, that attempted to take account of the intrinsic worth of emotional evaluations—might generate a richer conception of "justification." But why go to the trouble of developing such a theory of justification when doctrines informed by the evaluative conception of emotion can be applied in a satisfactory way without it?



sions. Voluntary manslaughter—which clearly does concern itself with the quality of the defendant's emotional motivations—thus falls into an explanatory blind spot for the dominant accounts of criminal law.

2. *The Model Penal Code Formulation.* — Under the Model Penal Code, an intentional homicide is graded as manslaughter when it “is committed under the influence of extreme mental or emotional disturbance for which there is reasonable explanation or excuse.”<sup>217</sup> The Code further provides that “[t]he reasonableness of such explanation or excuse shall be determined from the viewpoint of a person in the actor's situation under the circumstances as he believes them to be.”<sup>218</sup> In comparison with the common law formulation, the Model Penal Code version of voluntary manslaughter is decidedly mechanistic.

The Code is mechanistic at least partially by design. The drafters accepted the mechanistic premise that the impairment of self-control associated with emotion reduces both a person's culpability for bad acts and her deterrability.<sup>219</sup> Following through on this logic, they rejected as “arbitrary” the common law's requirements that the defendant be provoked and that she confine any retaliatory outburst to the source of the provocation.<sup>220</sup> According to the drafters, “the cause and intensity of the actor's emotion” might in some cases morally excuse him “even where he strikes out in a blinding rage and kills an innocent bystander.”<sup>221</sup>

In addition, the drafters “qualifie[d] the rigorous objectivity” of the common law.<sup>222</sup> The common law formulation measures the adequacy of a provocation according to its effect on a “reasonable person.”<sup>223</sup> This standard requires the jury to assess the gravity of the provocation against the background of contemporary community norms.<sup>224</sup> The Model Penal Code, in contrast, directs the jury to consider the “reasonableness” of the defendant's conduct “from the viewpoint of a person in *the actor's situation*.”<sup>225</sup> This formulation was intended to introduce a “larger element of subjectivity”<sup>226</sup> into the doctrine. Although “designedly ambiguous,” the term “situation” does not exclude a defendant's “exceptionally punctilious sense of personal honor,” her “abnormally fearful tempera-

---

217. Model Penal Code § 210.3(1)(b) (1980).

218. *Id.*

219. See *id.* § 210.3 cmt. 5(a) at 55–56.

220. See *id.* at 61.

221. *Id.*

222. *Id.* at 61–62.

223. See LaFave & Scott, *supra* note 145, § 7.10(b).

224. See *Director of Pub. Prosecutions v. Camplin*, 1978 App. Cas. 705 (appeal taken from Eng. C.A.); McAuley, *supra* note 203, at 146–48, 151–52.

225. Model Penal Code § 210.3(1)(b) (1980) (emphasis added); see *id.* cmt. 5 at 54 (Model Penal Code formulation “places far more emphasis than does the common law on the actor's subjective mental state”).

226. *Id.* cmt. 3 at 49.

ment," or other personal characteristics that "differentiate [her] from the hypothetical reasonable man" of the common law.<sup>227</sup>

Nonetheless, the drafters stopped short of making the Code formulation purely mechanistic. It is not enough under the Code for a defendant to show loss of control; there must still be a "reasonable explanation or excuse" for the defendant's "disturbance."<sup>228</sup> This does not necessarily require showing that the defendant's emotion was appropriate in light of community mores, but it does require showing that it satisfied some normative standard. "In the end," according to the drafters, "the question is whether the actor's loss of self-control can be understood in terms that arouse sympathy in the ordinary citizen."<sup>229</sup> It is plausible to think that juries will sometimes apply this constraint in a fashion that reflects the evaluative view.

But whatever conception of emotion the drafters might have meant the Code to embody, judicial interpretation has pushed it resolutely in a mechanistic direction. According to the courts, the Code formulation of manslaughter contemplates a degree of volitional impairment just short of insanity (which, under the Code, consists of a near-complete loss of self-control).<sup>230</sup> Thus, *any* affective experience sufficient to disable a person's "usual intellectual controls" or scramble "normal rational thinking" counts as an "extreme emotional disturbance."<sup>231</sup> To get to the jury, the defendant need show only that his feelings were sufficiently "intense," not that they were in any sense appropriate to his situation.<sup>232</sup>

Indeed, case law applying the Code is bristling with examples of defendants whose homicidal outbursts cannot be understood at all, much less understood as expressing appropriate judgments of value. A man apparently overwrought "by a combination of child custody problems, the inability to maintain a recently purchased home and an overwhelming fear of his brother" hunts his brother down and shoots him without provocation;<sup>233</sup> another who is high on drugs becomes extremely distressed when he perceives that a police officer is about to arrest his girlfriend for committing armed robbery;<sup>234</sup> and yet another stabs a woman in the throat and submerges her head in a bathtub for declining his offer of

---

227. Id. cmt. 5(a) at 62.

228. Id. § 210.3(1)(b).

229. Id. cmt. 5(a) at 63.

230. See *Patterson v. New York*, 432 U.S. 197, 206 (1977); *State v. Elliott*, 411 A.2d 3, 7 (Conn. 1979); *People v. Casassa*, 404 N.E.2d 1310, 1314 (N.Y. 1980).

231. *Elliott*, 411 A.2d at 8; see also *Casassa*, 404 N.E.2d at 1312 (woman telling obsessed defendant that she was not "falling in love" with him found sufficient to cause defendant extreme emotional disturbance, though court did not find reasonable explanation or excuse).

232. See *Elliott*, 411 A.2d at 8; *Casassa*, 404 N.E.2d at 1317.

233. *Elliott*, 411 A.2d at 5.

234. See *People v. Ford*, 423 N.Y.S.2d 402, 405 (1979).

liquor.<sup>235</sup> If the theme of the common law manslaughter cases is "virtuous rage," the theme of the Model Penal Code is "pathology."

This difference between the common law and Model Penal Code approaches informs not just the substance of the doctrines but the types of evidence that are important under each. Because of its evaluative underpinnings, the common law formulation has never been especially hospitable to psychiatric experts. The common law formulation does not insist that a defendant's passion be of any particular intensity, and in fact makes intensity immaterial in circumstances that fail to reveal the passion to be morally appropriate;<sup>236</sup> accordingly, in cases in which the evidence fails as a matter of law to show "adequate provocation" and lack of cooling time—issues governed by community norms—courts are likely to exclude expert psychiatric testimony altogether.<sup>237</sup> The Model Penal Code, in contrast, eliminates the doctrinal limitations designed to measure the appropriateness of a defendant's emotions. It thus puts a premium on the testimony of experts who can depict defendants' emotions in reductive, mechanistic terms.<sup>238</sup>

The career of the Model Penal Code formulation has not been a particularly happy one. Among the states that have adopted the Model Penal Code, relatively few enacted the Code version of voluntary manslaughter; moreover, a substantial number of the ones that did reverted to the common law formulation after only a short time.<sup>239</sup> Consumers of legal doctrine, at least, clearly prefer the evaluative position of the common law.

### C. *Premeditated Murder*

Most states distinguish not only between murder and manslaughter but also between different degrees of murder. Under the most popular formulation, first degree murder consists of all "premeditated" killings, and second degree of all other intentional killings (and some reckless ones, too).<sup>240</sup>

The prevailing accounts presuppose the mechanistic conception of emotion. One account is voluntarist:

Statutes . . . which distinguish deliberate and premeditated murder from other murder, reflect a belief that one who meditates an intent to kill and then deliberately executes it is . . . more culpable or less capable of reformation than one who kills on

---

235. See *Casassa*, 404 N.E.2d at 1312.

236. See *supra* text accompanying notes 154–163.

237. See, e.g., *Commonwealth v. Carr* 580 A.2d 1362, 1364 (Pa. Super. Ct. 1990); *State v. Gounagias* 153 P. 9, 13–15 (Wash. 1915).

238. See *Singer*, *supra* note 146, at 298–99.

239. See *Kadish & Schulhofer*, *supra* note 169, at 423.

240. See generally *LaFave & Scott*, *supra* note 145, § 7.7 (discussing degrees of murder).

sudden impulse . . . . The deliberate killer is guilty of first degree murder; the impulsive killer is not.<sup>241</sup>

This view equates culpability with the quality of the actor's volition and assumes that impassioned or impulsive behavior detracts from volition.

Another account reflects narrow consequentialism. According to this account, a person who dispassionately and methodically plans out a killing is more likely to succeed and to avoid detection than a person who kills suddenly in a fit of anger or fear. The premeditated killer is therefore more dangerous than other types of killers and should be punished more severely.<sup>242</sup> As with the mechanistic account of manslaughter, a killer's passion is regarded as a mitigating factor not because it expresses appropriate evaluations, but only because it supposedly reveals the actor to have a lesser propensity to produce undesired states of affairs.

The problem for these accounts, descriptively, is that they take "premeditation" literally. "Premeditation" is in fact one of the great fictions of the law. Many jurisdictions do not require any evidence of advance planning or thought to establish it; "no time is too short," they reason, "for a wicked man to frame in his mind his scheme of murder, and to contrive the means of accomplishing it."<sup>243</sup> Courts that adhere to this understanding have also consistently denied that "premeditation" is negated by evidence that a defendant killed on the basis of "a blind or irresistible impulse" or "uncontrolled emotions."<sup>244</sup> Consequently, in these jurisdictions at least, the law cannot be distinguishing between different classes of murderers on the grounds, or for the reasons, that the dominant mechanistic accounts suggest.

---

241. *Bullock v. United States*, 122 F.2d 213, 214 (D.C. Cir. 1941).

242. See, e.g., Posner, *An Economic Theory of the Criminal Law*, supra note 133, at 1222-23. It is sometimes said, in addition, that the impulsive or impassioned killer is less deterrable than the premeditated killer. See, e.g., *Bullock*, 122 F.2d at 214. Sophisticated consequentialists disavow this argument on the ground that deterrence considerations warrant more severe, not less severe, punishment for impassioned killers precisely because they have unusually strong desires to kill. See Posner, *An Economic Theory of the Criminal Law*, supra note 133, at 1223. It is unclear, however, why consequentialists do not view the intensity of the impassioned killer's desire to kill as a factor that justifies punishing her as severely as premeditated killers; in the absence of empirical evidence, how are we to know which effect dominates—the lower probability of detection and higher likelihood of success associated with planned killings or the greater desire to kill associated with impulsive or impassioned killings?

243. *Commonwealth v. Drum*, 58 Pa. 9, 16 (1868). See generally Joshua Dressler, *Understanding Criminal Law* 474-75 (2d ed. 1995) [hereinafter Dressler, *Understanding Criminal Law*].

244. *Commonwealth v. Carroll*, 194 A.2d 911, 918 (Pa. 1963) (quoting *Commonwealth v. Tyrrell*, 174 A.2d 852, 856-57 (Pa. 1961)); see *Commonwealth v. Weinstein*, 451 A.2d 1344, 1350 (Pa. 1982); *Jacobs v. Commonwealth*, 15 A. 465, 466 (Pa. 1888); see also *State v. Sikora*, 210 A.2d 193, 203 (N.J. 1965) ("A criminal act of that nature is nothing more than the consequence of an impulse that was not resisted.").

Although many commentators criticize it as arbitrary and unprincipled,<sup>245</sup> the fictional conception of premeditation is in fact perfectly coherent under the evaluative view. To begin, the evaluative view explains why killings that are premeditated in a literal sense are at least sometimes viewed as among the most heinous. When we say that a person “premeditated” we might mean that she killed another without passion—that the act of killing was “cold blooded.” Or, as in the case of someone who kills for money, we might have in mind that the person’s motive displays calculation and greed. Either way, it is the offender’s very willingness to kill when *not* motivated by a strong emotional valuation that expresses just how little regard she has for the victim’s life. The evaluative view, in other words, can give us reason to criticize a person for the emotions she *doesn’t* have as well as for the ones that she does have.

But it also isn’t surprising under the evaluative view that “premeditation” has not been confined to cold-blooded or calculated killings. From an evaluative perspective, it isn’t the case that *all* passionless killings are more heinous than *all* impassioned ones, because some emotional motivations for killing are themselves extraordinarily depraved. Imagine, for example, a man who on impulse sadistically tortures and kills a child.<sup>246</sup> Indeed, a view that categorically excluded all impassioned killings from first degree murder would proceed “imperceptibly to the absurd result that the more strange and brutal the act the more likely the actor is to be relieved of its criminal consequences.”<sup>247</sup> Because it is the *quality* of offenders’ motives (or lack thereof), and not the intensity or suddenness of them, that separates the most heinous murders from the rest, the evaluative view supports the predominant, fictional conception of “premeditation.”<sup>248</sup>

The career of “premeditation” is emblematic of the criminal law’s ambivalent stance toward emotions. The origins and no doubt the staying power of “premeditation” attest to the surface appeal of the mechanistic conception. But the subversion of it in most jurisdictions proves that commitment to the mechanistic view often runs only skin deep. Doctrines that start out mechanistic frequently metamorphosize into evaluative ones, which more completely account for decisionmakers’ judgments about the moral significance of emotions in assessing culpability. The result is obscure evaluation. Cardozo criticized the dominant conception

---

245. See, e.g., Dressler, *Understanding Criminal Law*, *supra* note 243, at 474–75 (suggesting that majority view undermines the principle “that a carefully considered and planned killing is worse than a spur-of-the-moment homicide”); LaFave & Scott, *supra* note 145, at § 7.7(a).

246. See 3 Sir James Fitzjames Stephen, *A History of the Criminal Law of England* 94 (1883).

247. *Weinstein*, 451 A.2d at 1349.

248. See Samuel H. Pillsbury, *Evil and the Law of Murder*, 24 U.C. Davis L. Rev. 437, 454 (1990) [hereinafter Pillsbury, *Evil and Murder*] (“We might see what courts have done with premeditation not so much as a subversion of rationality analysis, but as a covert move to find room for motivation analysis.”).

of premeditation on exactly this ground: the power to evaluate offenders' passions, he argued in a well-known appraisal of the doctrine, should be openly acknowledged and not shrouded "in a mystifying cloud of words."<sup>249</sup> In Part IV, we return to this theme, examining the real harm that this mystification imposes.

There remains at least one puzzle about this account of premeditation. The evaluative conception might explain why courts do not construe premeditation literally. But reduced to a fiction, the concept of premeditation does nothing on its own to identify the most reprehensible class of murders. If the law were really designed to implement the evaluative conception of emotion, wouldn't the division between first and second degree murder be marked by a rule (or set of rules) that expressly and concretely correlate the grading of a homicide to the quality of the emotions that move individuals to kill?

For two reasons, we believe the answer is "no," or at least "not necessarily." They are essentially the reasons that account for the contemporary reluctance of many jurisdictions to define "adequate provocation" as a matter of law. The first is the stubborn particularity of the evaluations that decisionmakers are likely to make in this setting. The emotions that inspire intentional killings are exceedingly heterogeneous and context-specific, and, as a result, so are the evaluations that observers make of them. A woman who coolly yet greedily plans the killing of her husband to obtain an inheritance seems more reprehensible than a woman who angrily and suddenly kills her husband upon discovering that he has abused their children; a man who methodically yet compassionately kills his terminally ill wife seems less reprehensible than a man who impulsively and sadistically shoots a stranger for a trivial offense.<sup>250</sup> Because any abstract set of rules is likely to track moral judgments very imperfectly, we might expect the law—if it is truly informed by evaluative content—to avoid mechanical definitions and instead use a fairly contentless standard (like "premeditation") as a placeholder for the sum total of the decisionmaker's moral intuitions, which can then be freely brought to bear on the facts of each particular case.<sup>251</sup>

The second reason that we might expect the law to avoid clear rules for distinguishing first from second degree murders has to do with how the power to make evaluative appraisals is allocated among different decisionmakers. Jurisdictions that employ premeditation as a fiction give the jury unbridled authority to determine what complexes of emotions, in-

---

249. Benjamin N. Cardozo, *What Medicine Can Do for Law* 27 (1930).

250. See generally Model Penal Code § 210.6 cmt. 3(b) at 127 (1980) (noting unfairness of reducing all impassioned killings to manslaughter).

251. See generally Dan M. Kahan, *Lenity and Federal Common Law Crimes*, 1994 Sup. Ct. Rev. 345, 399–418 (defending underspecified statutory definitions for certain crimes that take diverse forms).

tentions, and other mental states count as "premeditation";<sup>252</sup> jurisdictions that give more specific content to "premeditation" invest more power in the court, which can use the legal definition of the term to second guess the jury's verdict.<sup>253</sup> It is perfectly consistent with the evaluative view to prefer the jury's judgment to the court's on the ground that the jury has superior access to community mores, which form the background against which a defendant's emotions are to be assessed.<sup>254</sup>

But while we believe for these reasons that the fictional character of premeditation is consistent with the evaluative view, we do not believe that the evaluative approach compels that the grading of intentional murders be completely unspecified and discretionary. The law need not always passively reflect community norms when assessing offenders' emotions; it can also check, constrain, and critically shape those norms. This is exactly what happens, for example, when courts determine that some acts are inadequate provocation as a matter of law for purposes of manslaughter doctrine.<sup>255</sup> Similarly, without attempting fully to specify which types of murders are the most heinous, legislatures (and possibly courts)<sup>256</sup> could provide that some murders—reflecting especially heinous emotional motivation—must necessarily be treated as first degree.<sup>257</sup> We defend such an approach on normative grounds in Part IV.<sup>258</sup>

#### D. *Self-Defense*

The law permits private citizens to use deadly force under certain circumstances. The circumstances vary across jurisdictions. All states permit a person to use deadly force when it reasonably appears necessary to protect herself from an imminent, unlawful, and unprovoked threat to her own life; but some also permit the use of deadly force to protect life even when use of deadly force appears unnecessary, or to protect inter-

---

252. See, e.g., *United States v. Brown*, 518 F.2d 821, 825 (7th Cir. 1975) ("It is well settled that the question of whether or not reflection and consideration amounting to deliberation required for first degree murder actually occurred must be determined by the jury, properly instructed by the court, from the facts and circumstances of the case."); *Commonwealth v. Drum*, 58 Pa. 9, 16 (1868) ("The law fixes upon no length of time as necessary to form the intention to kill, but leaves the existence of a fully formed intent as a fact to be determined by the jury, from all the facts and circumstances in the evidence.").

253. See, e.g., *People v. Anderson*, 447 P.2d 942, 949 (Cal. 1968) (outlining factors that guide appellate court's scrutiny of jury verdict of first degree murder); *People v. Morrin*, 187 N.W.2d 434, 452 (Mich. Ct. App. 1971) (applying similarly exacting review).

254. See *supra* text accompanying notes 105–115.

255. See *supra* text accompanying notes 150–180.

256. Cf. *State v. Jones*, 257 Kan. 856, 872–73 (1995) (in prosecution for first-degree murder, trial court can refuse to instruct on second-degree if evidence is legally insufficient to support lesser charge); *State v. Walls*, 463 S.E.2d 738, 762 (N.C. 1995) (same); *State v. Sexton*, 724 S.W.2d 371, 374 (Tenn. Crim. App. 1986) (same).

257. See, e.g., *Pillsbury, Evil and Murder*, *supra* note 248, at 480–81 (attempting a fuller specification of aggravated murder).

258. See *infra* text accompanying note 420.

ests besides life.<sup>259</sup> The appropriate scope of the doctrine is an issue of ongoing dispute. Without an appreciation of how the mechanistic and evaluative conceptions of emotion have shaped the doctrine, it is impossible fully to understand what is at stake in this debate.

It might be thought that self-defense is either unconcerned with emotions or concerned with them only incidentally. Self-defense is a paradigmatic justification,<sup>260</sup> and justifications are typically understood to refer to acts that produce preferred states of affairs.<sup>261</sup> Accordingly, the law of self-defense can be viewed as embodying normative judgments about when the death of an aggressor is preferable to the destruction of the interest that the aggressor threatens.<sup>262</sup> The emotions of the person who justifiably uses self-defense, it might be inferred, are simply immaterial.

But this conclusion would be wrong. It is impossible to understand *why* self-defense doctrine prefers the death of the aggressor to the impairment of certain interests without appreciating the law's assessment of the defendant's emotions. To be sure, the defense doesn't require proof that a particular defendant was motivated by a particular emotion, but the contours of the doctrine nevertheless reflect understandings about what kinds of emotions a "reasonable person"—that is, a person with ordinary or appropriate sensibilities—would experience in particular situations.

The classic account of self-defense doctrine assumes a mechanistic conception of emotion. Because "law . . . respects the passions of the human mind," Blackstone explained, it permits the man confronted with "external violence . . . to do himself that immediate justice, to which he is prompted by nature, and which no prudential motives are strong enough to restrain."<sup>263</sup> This account of self-defense has both voluntarist and consequentialist underpinnings: if a person has no realistic choice but to use deadly force, then her use of such force is neither culpable nor deterrable.

---

259. See LaFave & Scott, *supra* note 145, § 5.7.

260. See *id.* § 5.7 & n.1; Robinson, *Systematic Analysis*, *supra* note 205, at 236.

261. See *supra* text accompanying notes 205–208.

262. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 859; Robinson, *Systematic Analysis*, *supra* note 205, at 236.

263. 3 William Blackstone, *Commentaries* \*3–4; see also Thomas Hobbes, *Leviathan* 345–46 (Penguin Classics 1985) (1651) ("If a man by the terror of present death, be compelled to doe a fact against the Law, he is totally Excused; because no Law can oblige a man to abandon his own preservation. . . . Nature . . . compels him to the fact."); Oliver W. Holmes, Jr., *The Common Law* 47 (Dover 1991) (1881) ("[T]he law cannot prevent [use of deadly force in self-defense] by punishment, because a threat of death at some future time can never be a sufficiently powerful motive to make a man choose death now in order to avoid the threat."). Blackstone's position may have been less mechanistic than it appears; to say that a person "prompted by nature" to engage in a particular act is impelled to that act by unreasoning forces outside her control is to assume a contentious view of nature that Blackstone himself may not have held. Nevertheless, Blackstone's position has been perceived in a mechanistic light and has influenced the doctrine accordingly. See Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 856–57.



The mechanistic understanding supports a narrow formulation of self-defense doctrine. Under this reasoning, the defense must be confined to circumstances of "utmost . . . necessity," for only then can the "primal impulse" of "self-preservation" be understood to overwhelm a person's capacity to refrain from the use of deadly violence.<sup>264</sup>

Another account of self-defense doctrine assumes an evaluative conception of emotion. Under this view, the law accommodates not just the fear but also the pride of a person confronted by wrongful aggression. The broader contours of the doctrine reflect the extent to which the law is prepared to endorse valuations of honor or dignity in circumstances in which these goods can be protected only by deadly force.

This evaluative understanding is at work, for example, in departures from the "necessity" requirement. The majority of American jurisdictions take the position that an individual is entitled to repel a life-threatening assault with deadly force even if she could safely evade her attacker.<sup>265</sup> Although it is possible to account for this rule in mechanistic terms,<sup>266</sup> the nineteenth-century courts that fashioned it clearly had the defendant's honor, not his fear, in mind.<sup>267</sup> It would be demeaning, they reasoned, to require " 'a true man . . . to fly from an assailant, who by violence or surprise maliciously seeks to take his life or do him enormous bodily harm.' " <sup>268</sup> The English common law requirement that a man "retreat to the wall" before using deadly force was contrary to " 'the tendency of the American mind,' " which rebelled at the suggestion that " 'a person, being without fault, and in a place where he has a right to be,' " must " 'flee when assailed, to avoid chastisement, or even to save human life.' " <sup>269</sup> Recent empirical studies suggest that this sensibility remains widespread.<sup>270</sup>

In addition, many jurisdictions that do impose a general duty of retreat nonetheless hold that a person may stand her ground when confronted by an attacker in her own residence.<sup>271</sup> Insofar as the law in these jurisdictions does not similarly excuse the obligation to avoid con-

---

264. *State v. Norman*, 378 S.E.2d 8, 12-13 (N.C. 1989); see also *United States v. Peterson*, 483 F.2d 1222, 1229-30 (D.C. Cir. 1973) (self-defense is "[h]inged on the exigencies of self-preservation" where the defender believed that "his response was necessary to save himself" from imminent peril of death or serious bodily harm).

265. See LaFave & Scott, *supra* note 145, § 7.10(f).

266. Justice Holmes did, explaining that "detached reflection cannot be demanded in the presence of an uplifted knife." *Brown v. United States*, 256 U.S. 335, 343 (1921).

267. See LaFave & Scott, *supra* note 145, § 5.7(f) (noting that "no retreat" doctrine reflects "a policy against making one act a cowardly and humiliating role").

268. *Beard v. United States*, 158 U.S. 550, 561 (1895) (Harlan, J.) (quoting *Erwin v. State*, 29 Ohio St. 186, 193, 199 (1876)).

269. *Id.* at 561-62 (quoting *Runyan v. State*, 57 Ind. 80, 84 (1877)).

270. See Robinson & Darley, *supra* note 177, at 60.

271. See LaFave & Scott, *supra* note 145, § 5.7(f).

frontation when a person is attacked in another's home,<sup>272</sup> it is implausible to view this rule as reflecting a mechanistic presumption that persons are incapable of voluntarily choosing flight when attacked in close quarters. Again, the better explanation is honor: the so-called "castle" doctrine is understood to spare an individual the indignity of being made "a fugitive from his own home."<sup>273</sup>

The evaluative conception is also at work when the law enlarges the class of interests that may be vindicated by use of deadly force. In some jurisdictions, a person can use deadly force to repel not only life-threatening attacks, but also a host of non-life-threatening invasions, including robberies and sexual assaults.<sup>274</sup> However plausible the mechanistic view might be where a person's "self-preservation" is at stake, it seems quite unpersuasive to suggest that human beings have a "primal impulse" to kill non-deadly muggers or even non-deadly rapists. The better understanding is that the law (in some states) is prepared to endorse the valuation of dignity and honor expressed in a person's refusal to submit to certain wrongful acts that powerfully convey that person's subordination to the will of another.<sup>275</sup>

And even when the law is not prepared to do this, juries often are. The evaluative conception is easy to spot in verdicts that flagrantly disregard the constraints of the classic self-defense doctrine.<sup>276</sup> Historically, findings of self-defense were one device by which juries enforced the "unwritten rule" that men may justifiably kill their wives' paramours.<sup>277</sup> The acquittal of Bernhard Goetz is a contemporary example—a singularly ugly one, given the racial dimension of the case—of how juries can use self-defense to endorse the valuation of honor and dignity expressed

---

272. See, e.g., *DeVaughn v. State*, 194 A.2d 109, 112 (Md. 1963) (distinguishing the self-defense law on the basis that "the house where the shooting occurred was not [the defendant's] home or 'castle'").

273. *People v. Tomlins*, 107 N.E. 496, 497 (N.Y. 1914) (Cardozo, J.).

274. See, e.g., N.H. Rev. Stat. Ann. § 627:4(II)(b)(c) (1986); N.Y. Penal Law § 35.15 (McKinney 1987); Tex. Penal Code Ann. § 9.32 (West 1994).

275. See, e.g., *State v. Philbrick*, 402 A.2d 59, 62–63 (Me. 1979) (statutory privilege to use deadly force to prevent nonconsensual touching of a person's genitals "reflect[s] a societal decision that the public's interest in being free from the threat of [certain] crimes . . . outweigh[s] the high value society otherwise places on human life"); *Moore v. State*, 237 S.W. 931 (Tex. 1922) (holding that parent was justified in using deadly force to prevent statutory rape of seventeen-year-old daughter).

276. See generally Kalven & Zeisel, *supra* note 177, at 221–41 (documenting various settings in which jury implicitly condones excessive use of force in response to insulting or demeaning behavior); Robinson & Darley, *supra* note 177, at 56–57 (documenting common sensibility that downgrades culpability of person who uses deadly force to repel offensive but non-life-threatening harassment).

277. See Jeremy D. Weinstein, Note, *Adultery, Law, and the State: A History*, 38 *Hastings L.J.* 195, 229 (1986); see also Kalven & Zeisel, *supra* note 177, at 234–36 (documenting jury sympathy toward defendants who use violence in response to infidelity).

when an enraged person uses deadly force to avenge non-life-threatening transgressions.<sup>278</sup>

So far, we have attempted to illustrate the influence of the evaluative conception of emotion by showing how it has informed broad formulations of the self-defense doctrine. But there is no necessary connection between the evaluative view and broad formulations; the law could as easily defend narrow formulations of self-defense to criticize excessive valuations of honor and dignity.

Indeed, one reaction against a broadly defined self-defense privilege can itself be explained in evaluative terms. In a turn-of-the-century article, Joseph Beale argued that the conception of honor that informs the "true man" doctrine ought to be condemned, not endorsed:

The feeling at the bottom of the [the rule] is one beyond all law; it is the feeling which is responsible for the duel, for war, for lynching; the feeling which leads a jury to acquit the slayer of his wife's paramour; the feeling which would compel a true man to kill the ravisher of his daughter. We have outlived dueling, and we deprecate war and lynching; but it is only because the advance of civilization and culture has led us to control our feelings by our will.<sup>279</sup>

Beale's critique of the "true man" doctrine is evaluative because it assumes that "feelings" reflect judgments of value that are themselves open to moral assessment. One might understand Beale to be attacking emotional valuations across the board: law contributes to the "advance of civilization" by seeking to subdue our "feelings," which invariably attribute inappropriate value to goods like esteem. But we think that Beale can also be understood in more Aristotelian terms:

A really honorable man, a man of truly refined and elevated feeling, would perhaps always regret the apparent cowardice of a retreat, but he would regret ten times more, after the excitement of the contest was past, the thought that he had the blood of a fellow-being on his hands. It is undoubtedly distasteful to retreat; but it is ten times more distasteful to kill.<sup>280</sup>

This argument asserts not that law should be morally suspicious of *all* emotions, but rather that law should prefer emotions that express morally true valuations to emotions that express morally false ones. The law of self-defense should be defined, Beale argues, so as to endorse (and propagate) a "truly refined and elevated feeling" that expresses an appropriately high valuation of the worth of all persons, even those who act wrongly.

---

278. See generally George P. Fletcher, *A Crime of Self-Defense: Bernhard Goetz and the Law on Trial* 179 (1988) [hereinafter Fletcher, *Self-Defense*] (suggesting that the issue was "the right of decent citizens to hold their ground against the terrorizing effect of the mugging subculture").

279. Joseph H. Beale, Jr., *Retreat from a Murderous Assault*, 16 Harv. L. Rev. 567, 581 (1903).

280. *Id.*

The terms of Beale's critique of the "true man" rule have proven at least modestly influential.<sup>281</sup> And where courts reject the rule on Beale's grounds, they are clearly adopting an evaluative stance.

Uncovering the influence of the mechanistic and evaluative conceptions of emotion illuminates contemporary as well as historical disputes over the contours of self-defense doctrine. These conceptions of emotions figure prominently, for example, in debates about whether the doctrine should be formally modified to permit domestic abuse victims to use deadly force in the absence of immediately life-threatening behavior. One argument for enlarging the doctrine to cover such cases is mechanistic: it asserts that the debilitating influence of such abuse on a person's capacity to appraise her situation and to exercise self-control should be integrated into the law's understanding of when the use of deadly force is objectively reasonable.<sup>282</sup> Another is evaluative. According to this account, the requirement of an "imminent" deadly threat should be relaxed not because persistent abuse effectively destroys a person's agency, but because demanding that such a person forgo violence is to insist that she endure an experience of degradation and humiliation inconsistent with human flourishing.<sup>283</sup> An evaluative account of the battered woman's fear also explains why acknowledging *her* self-defense claim does not entail acknowledging the claims of offenders who, like Goetz, are impelled to violence by intense but morally inappropriate emotions.

Neither account has proven particularly persuasive in American jurisdictions. The relevance of evidence of long-term abuse, and even expert testimony relating to "battered woman syndrome" and like conditions, is now widely accepted.<sup>284</sup> But courts have declined to relax the imminence requirement or make other formal adjustments in the doctrine on the ground that the defense must be strictly confined to circumstances in which deadly force is objectively necessary to protect the life of the defendant.<sup>285</sup>

---

281. See, e.g., *State v. Abbott*, 174 A.2d 881, 884 (N.J. 1961):

[T]he advocates of no-retreat say the manly thing is to hold one's ground, and hence society should not demand what smacks of cowardice. Adherents of the retreat rule reply it is better that the assailed shall retreat than that the life of another be needlessly spent. They add that not only do right-thinking men agree, but further a rule so requiring may well induce others to adhere to that worthy standard of behavior.

282. See Lenore E. Walker, *The Battered Woman Syndrome* 142-43 (1984).

283. See, e.g., Phyllis L. Crocker, *The Meaning of Equality for Battered Women Who Kill Men in Self-Defense*, 8 Harv. Women's L.J. 121, 130-31 (1985); Charles P. Ewing, *Psychological Self-Defense*, 14 Law & Hum. Behav. 579, 586-87 (1990); Martha R. Mahoney, *Legal Images of Battered Women: Redefining the Issue of Separation*, 90 Mich. L. Rev. 1, 2 (1991).

284. See Kadish & Schulhofer, *supra* note 169, at 819.

285. See, e.g., *State v. Norman*, 378 S.E.2d. 8, 12 (N.C. 1989) (denying jury instruction on self-defense since no evidence was introduced tending to show "reasonable" belief of necessity, despite evidence introduced relating to battered wife syndrome);

Our analysis suggests that this response is inadequate, at least in jurisdictions that employ the "true man" or "castle" doctrines or that permit the use of deadly force to protect interests other than the defendant's life. These rules already depart from the rule of "objective necessity," and they do so on evaluative grounds. Perhaps there is a principled basis for simultaneously endorsing the emotional motivation of a person who kills rather than endure the indignity of seeking escape and criticizing the emotional motivation of a woman who kills rather than endure the degradation of continued abuse. But to say that the law cannot be adjusted to privilege deadly force in the latter context because the law *never* ranks the defendant's honor and dignity over a wrongful aggressor's life betrays either extreme confusion or hypocrisy.<sup>286</sup>

#### E. Duress

Imagine a woman who agrees to help commit an armed robbery after a man threatens to injure her or her children if she refuses. She can assert duress,<sup>287</sup> which relieves a person from liability for acts that the person reasonably believes to be necessary to avert threatened bodily harm or death.<sup>288</sup>

The rationale of duress is contested. It is often said to be "a paradigmatic example of an excuse":<sup>289</sup> the woman in our hypothetical is afforded a defense not because engaging in armed robbery is the morally appropriate thing to do—that crime may itself expose numerous innocent persons to risk—but because the man's threat renders the woman blameless for her acts.<sup>290</sup> But the defense is sometimes understood as a justification; the woman in our example should have a defense, on this view, only if the threatened harm to her or her children constitutes a "greater evil" than her participation in the robbery.<sup>291</sup>

The source of this and other disagreements concerning duress can again be attributed to a failure to recognize the influence of the evaluative conception of emotion. Once the role of the evaluative view is identi-

---

Richard A. Rosen, On Self-Defense, Imminence, and Women Who Kill Their Batterers, 71 N.C. L. Rev. 371 (1993).

286. See Rosen, *supra* note 285, at 390, 396–97.

287. Cf. *People v. Romero*, 13 Cal. Rptr. 2d 332, 340 (Cal. Ct. App. 1992) (holding that there was "a reasonable probability" that expert testimony on "Battered Wife Syndrome" would have persuaded jury to accept duress defense); *Morrison v. State*, 546 So. 2d 102, 103 (Fla. Dist. Ct. App. 1989).

288. See generally LaFave & Scott, *supra* note 145, § 5.3. The precise contours of the doctrine vary across jurisdictions. In some states, the threatened harm must be potentially deadly, in others not; some jurisdictions allow the defense in homicide cases, but most do not. See Kadish & Schulhofer, *supra* note 169, at 911.

289. Fletcher, *Rethinking Criminal Law*, *supra* note 204, at 830.

290. See Joshua Dressler, *Exegesis of the Law of Duress: Justifying the Excuse and Searching for Its Proper Limits*, 62 S. Cal. L. Rev. 1331, 1373 (1989) [hereinafter Dressler, *Exegesis*]; Robinson, *Systematic Analysis*, *supra* note 205, at 221–22.

291. See LaFave & Scott, *supra* note 145, § 5.3(a).

fied, moreover, it appears that duress, like voluntary manslaughter, should be characterized as neither a justification nor an excuse as those concepts are conventionally understood.<sup>292</sup>

Mechanistic accounts of duress are commonplace. The most popular account depicts the defense in voluntarist terms: duress excuses because (and when) "the will of the accused has been overborne by threats . . . so that the commission of the alleged offence was no longer the voluntary act of the accused."<sup>293</sup> Consistent with the mechanistic conception of emotion, this account sees intense fear as exculpating because it impairs "a person's capacity to choose to comply with the law."<sup>294</sup> Another account reflects the narrow form of consequentialism popular among criminal law theorists: duress excuses because (and when) the threat of harm to the offender negates any inference that she harbors a general propensity to commit crimes.<sup>295</sup> What makes this account mechanistic is that it views the offender's fear as revealing her disposition to promote or frustrate desired states of affairs, not as embodying valuations that are themselves subject to moral assessment.

The voluntarist account is clearly inadequate as a descriptive matter. Doctrinally, the experience of fear is not sufficient to support a claim of duress; the fear must also be objectively reasonable.<sup>296</sup> Thus, if the woman in our example was threatened only with a minor battery (say, a slap in the face), she would not have a defense—no matter how frightened she became—for "a person of reasonable firmness" would have resisted.<sup>297</sup> This objective standard assumes that a person can legitimately

---

292. For a penetrating analysis of duress that reaches a similar conclusion, see Claire O. Finkelstein, *Duress: A Philosophical Account of the Defense in Law*, 37 *Ariz. L. Rev.* 251 (1995).

293. *Regina v. Hudson*, [1971] 2 All E.R. 244, 246; see also *People v. Luther*, 232 N.W.2d 184, 186–87 (Mich. 1975) ("A successful duress defense excuses the defendant from criminal responsibility for an otherwise criminal act because the defendant was compelled to commit the act; the compulsion or duress overcomes the defendant's free will . . .").

294. Peter W. Low et al., *Criminal Law: Cases and Materials* 613 (2d ed. 1986); see also Robinson, *Systematic Analysis*, *supra* note 205, at 225 (duress excuses because "actor lacks the capacity to control his conduct").

295. See Brandt, *supra* note 136, at 174–75, 182, 190–91. Another consequentialist theory explains duress on the ground that it is pointless to punish persons moved to criminality by threats, since they are essentially undeterrable. See, e.g., Glanville Williams, *Criminal Law: The General Part* 756 (2d ed. 1961); Steven Shavell, *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, 85 *Colum. L. Rev.* 1232, 1257 (1985). But this argument seems unpersuasive. Presumably, even those faced with serious threats take into account the potential consequences of breaking the law; accordingly, from a deterrence point of view, it would make more sense to increase punishment for such persons to counteract the effects of the threat on their motivation to commit crimes. Cf. 2 Stephen, *supra* note 246, at 107 ("[I]t is at the moment when temptation to crime is strongest that the law should speak most clearly and emphatically to the contrary.").

296. See generally LaFave & Scott, *supra* note 145, § 5.3.

297. See, e.g., Model Penal Code § 2.09(1) (1985); *Regina v. Graham*, [1982] 1 W.L.R. 294, 300; see also *Knight v. State*, 601 So. 2d 403, 405 (Miss. 1992) ("[W]here a

be held accountable for acting on morally inappropriate fears, however intense. An account that stresses volitional impairment cannot explain this normative limitation.<sup>298</sup>

A narrow consequentialist account comes closer to justifying this aspect of the doctrine. The consequentialist view accepts that offenders who experience fear too readily should be punished; individuals who lack reasonable firmness are sufficiently susceptible to manipulation, and hence sufficiently dangerous to the community, to warrant incapacitation. The same cannot necessarily be said, however, about an individual who commits a crime only in response to a serious threat—one that even a person of reasonable firmness would have been unable to resist. Because there is no reason to infer that such an individual poses a higher-than-average risk of danger to the community, punishing her would be a waste.<sup>299</sup>

But the mechanistic underpinnings of this account deprive it of sufficient explanatory force. Cases applying duress frequently use a subtle, qualitative standard to appraise offenders' fears. For example, whereas a woman who commits armed robbery to avoid threatened bodily harm to herself is likely to have a defense, a mother who fails to protect her child from abuse in order to avert the same harm is likely *not* to have one.<sup>300</sup> These results seem mysterious if, following the consequentialists, we equate "reasonable firmness" only with an appropriately high level of tolerance for personal suffering; from a consequentialist point of view, the two women seem equally moved to criminality by the threat of bodily harm. The distinction between the two cases is much easier to explain, however, if we assume that duress embodies an evaluative conception of emotion: the fear of the first woman expresses an acceptable valuation of her own welfare relative to that of strangers, while that of the second shows that she values her own welfare excessively relative to that of her child.

It might be possible to reconcile the consequentialist account with the case law by specifying more precisely the state of affairs that the law is trying to maximize. The results in our two cases suggest that legal decisionmakers prefer the woman's participation in armed robbery to her submission to a beating, but prefer her submission to a beating to her acquiescence in the abuse of her child. Accordingly, it remains consistent with a consequentialist strategy to exculpate the first woman but not

---

person reasonably believes that he is in danger of physical harm he may be excused for some conduct which ordinarily would be criminal.").

298. See Kadish, *Excusing Crime*, *supra* note 129, at 272–75 (acknowledging gap between scope of duress defense and voluntarist principle of moral accountability).

299. See Brandt, *supra* note 136, at 182.

300. Compare *People v. Romero*, 13 Cal. Rptr. 2d 332, 340 (Cal. Ct. App. 1992), and *Morrison v. State*, 546 So. 2d 102, 103 (Fla. Dist. Ct. App. 1989) with *United States v. Webb*, 747 F.2d 278, 283 (5th Cir. 1984), cert. denied, 469 U.S. 1226 (1985) and *State v. Lucero*, 647 P.2d 406 (N.M. 1982).

the second based on what their fears reveal about their respective propensities to engage in socially undesirable conduct.

But once reformulated in this fashion, the consequentialist position is no longer sufficiently narrow. For the only basis that legal decisionmakers have for reaching different results in our two cases is that they are prepared to endorse the first woman's valuation of her own welfare but not the second woman's. As with voluntary manslaughter, the consequentialist theory has descriptive purchase only when it uses the valuations internal to the evaluative conception of emotion to identify preferred states of affairs. This is indeed a form of consequentialism—one that we believe is quite justifiable<sup>301</sup>—but it is too rich to be accounted for under the wealth-maximizing theory of value that narrow consequentialists employ.

Although necessarily diverse, the emotional evaluations that inform duress doctrine share a theme: the legitimate love of one's own. Typically, a defendant asserts the defense of duress when the social losses associated with her act exceed (in some sense) the threatened loss to her or her family members.<sup>302</sup> Such behavior is tolerated by the law not just because it is inevitable that an individual will prefer herself and her loved ones to the public at large, but also because it is at least sometimes morally appropriate to have such a preference. Intimate relationships are defined, ultimately, by intense partiality; we cannot at the same time affirm the moral validity of such relationships and condemn someone for fears that reveal that she values the well-being of family and friends more than she values the well-being of strangers. But clearly one's love of one's own faces moral limits; complex social norms define when such a preference is legitimate and when it is not. "Duress" captures the interplay of at least some of these norms and helps to regulate the interaction between the privilege to love one's own and the duty to treat all persons with concern and respect.

This evaluative interpretation challenges not only the dominant mechanistic accounts, but also the conception of the defense as an *excuse*. According to the evaluative view, duress exculpates not because (and when) threats vitiate a person's moral agency, but because (and when) a person's fear expresses a rational and morally appropriate assessment of her circumstances.<sup>303</sup> The woman who aids and abets an armed robbery to avoid bodily harm to herself or her family expresses appropriate love

---

301. See *infra* text accompanying notes 388–89.

302. If it were clear that her conduct enhanced net welfare, then she could likely assert the "necessity" or "lesser evil" defense. See Robinson, *Systematic Analysis*, *supra* note 205, at 213–14; see also Model Penal Code § 3.02(1)(a) (1985) (criminal act is justified where "the harm or evil sought to be avoided by such conduct is greater than that sought to be prevented by the law defining the offense").

303. See, e.g., *Knight v. State*, 601 So. 2d 403, 407 (Miss. 1992) (reversing for exclusion of evidence that "bears on the rationality of Knight's fear under the circumstances").



of her own; the woman who tolerates child abuse to avoid such harm loves herself too much and her child too little. Because duress exculpates only when a person's emotional motivation expresses morally appropriate valuations, it seems obtuse to insist that the defense presupposes that the actor has done the "wrong" thing. Indeed, where the defense is recognized, it will often seem that the actor did the only morally right thing *for her*. Were a person to expose her family to great harm rather than comply with a demand that she participate in a fraud, for example, she would likely be regarded as a monster, not a hero.<sup>304</sup>

This is not to say that duress, as we've explained it, should be viewed as a "justification." Consider again the woman who agrees to assist an armed robbery in response to coercive threats. She may be morally entitled (perhaps even morally obliged) to prefer her or her family's welfare to that of strangers, and may thus have a defense of duress. But given the risk that her actions create for innocent third parties, it cannot necessarily be said, from a consequentialist point of view, that her participation in the robbery results in a preferred state of affairs. In addition, the social norms that define the legitimate love of one's own are agent relative.<sup>305</sup> To protect her child from a threatened harm, a mother might be warranted in participating in a crime that exposes many others to harm; but it is unlikely that someone with no special relationship with the child would be warranted in committing such a crime.<sup>306</sup> Similarly, to avoid a threat of harm to himself and his family, a private citizen who collaborates with the enemy abroad might be able to assert a defense of duress;<sup>307</sup> but a soldier, even if imprisoned by the enemy and threatened with deadly reprisal, almost certainly cannot because he has a special obligation to prefer the welfare of his countrymen and fellow soldiers to his

---

304. Cf. *United States v. Garner*, 529 F.2d 962, 969-70 (6th Cir. 1976) (refusal to instruct jury on defendant's claim that she acted under duress of anonymous threats on her daughter's life was reversible error); *State v. Toscano*, 378 A.2d 755, 761-62 (N.J. 1977) (harm threatened need not be directed at actor, but can be directed at close friend or relative).

305. See Dressler, *Exegesis*, *supra* note 290, at 1354. See generally Anderson, *supra* note 191, at 73-79 (discussing agent-relative valuations); Thomas Nagel, *The Limits of Objectivity*, in 1 *Tanner Lectures on Human Values* 77, 102-03 (Sterling M. McMurrin ed., 1980) (same).

306. Douglas N. Husak, *Justifications and the Criminal Liability of Accessories*, 80 *J. Crim. L. & Criminology* 491, 511-12 (1989) (claiming that only someone with "special relationship" to another person can prefer that person's welfare to the welfare of multiple other persons).

307. Cf. *Kawakita v. United States*, 343 U.S. 717, 734-37 (1952) ("An American . . . who is charged with playing the role of traitor may defend by showing that force or coercion compelled such conduct"); *Rex v. Steane*, [1947] 1 All E.R. 813 (allowing duress defense for British subject charged with assisting the German army, where the defendant claimed he did so for the safety of his wife and children).

own.<sup>308</sup> The sensitivity of duress doctrine to an actor's social role defies the "universality" criterion for justifications.<sup>309</sup>

If this lack of fit is a problem, the difficulty lies with the categories of "justification" and "excuse" and not with the evaluative conception of duress. These concepts, as we've explained, make no space for the role that emotional evaluations play in moral appraisals.<sup>310</sup> As a result, they necessarily fail to account for doctrines such as duress, in which emotional evaluations are central.

#### F. *Voluntary Act*

It is axiomatic that criminal liability presupposes a "voluntary act."<sup>311</sup> It might be thought that this doctrine confounds our account. We have argued that criminal law doctrines frequently assess the quality of actors' emotional motivations, and that they do so on the assumption that offenders are responsible for their characters. The commission of a crime, however, is not the only evidence, and probably not even the best evidence, that a person's character is defective. Yet the law steadfastly refuses to punish mere "status" crimes.<sup>312</sup> Does this show that the criminal law is in fact antagonistic to the basic premises of the evaluative view?<sup>313</sup> We believe that the answer to this question is no; indeed, we will argue that the evaluative view is essential to make sense of what the law counts as a "voluntary act."

It is useful to disaggregate the requirement that there be an "act" from the requirement that it be "voluntary." The "act" requirement poses no difficulty for the evaluative view as we've defined it. Our core descriptive claim is a modest one: the moral quality of an offender's emotional motivations is a relevant consideration in criminal law. This claim is perfectly compatible with a wide range of rationales for insisting that crimi-

308. Cf. *United States v. Fleming*, 7 C.M.A. 543, 23 C.M.R. 7 (1957) (American soldier's fear of torture and possible death in prison camp insufficient as defense to charge of collaborating with the enemy).

309. See Alan Wertheimer, *Coercion* 167-68 (1987); Finkelstein, *supra* note 292, at 280-82. Again, it is possible to develop an enriched form of consequentialism that takes an actor's role into account when identifying the preferred state of affairs. See, e.g., Sen, *Rights and Agency*, *supra* note 134, at 29-30. However, such agent-relative evaluations are likely to be exceedingly diverse and particular; it thus seems unlikely that this richer form of consequentialism could be usefully translated into any abstract theory of justification. But cf. Dressler, *Exegesis*, *supra* note 290, at 1354-55 (suggesting that the theory of justification can accommodate agent-relative valuation). Indeed, it's unclear why any such theory would even be necessary, assuming that agent-centered evaluations (including those embodied in an actor's emotions) can generate satisfactory results in cases as they arise.

310. See *supra* text accompanying notes 203-216.

311. See, e.g., LaFare & Scott, *supra* note 145, at § 3.2(c).

312. Indeed, such a stance is constitutionally compelled. See *Robinson v. California*, 370 U.S. 660, 667 (1962).

313. For critiques of so-called "character theories" of criminal law, see R.A. Duff, *Choice, Character, and Criminal Liability*, 12 *Law & Phil.* 345, 371-80 (1993); Moore, *Choice*, *supra* note 129, at 48-49, 54-55.

nal liability be confined to acts; all that it presupposes is that when criminal law purports to regulate behavior, the doctrine is structured in a way that permits the quality of a defendant's emotions to be taken into account. If anything, the evaluative view presupposes acts that are capable of expressing the evaluations that emotions embody.<sup>314</sup> So, however embarrassing the act requirement may be to some "character" theories of criminal law,<sup>315</sup> it causes no embarrassment to ours.

To be sure, our core descriptive claim does focus on "character," but in a fairly limited and nontheoretical way. The evaluative view asserts that it is appropriate to take the emotions of a criminal defendant into consideration—whether or not they are willed—because she remains accountable for being the kind of person who experiences such emotions. This claim identifies responsibility for character as a sufficient ground for holding persons accountable for some acts that are not wholly attributable to individual choice.<sup>316</sup> But the evaluative view does not require that the criminal law or any other body of law take a more thoroughgoing interest in the state of citizens' characters. Indeed, the evaluative view, as we have developed it, does not even depend on strong theoretical claims about what "character" is apart from the emotions that individuals actually experience.

This is not to say that even this limited use of character is uncontentious. In particular, we think the claim that individuals are necessarily or always responsible for their emotional lives is clearly false.<sup>317</sup> Consequently, to be just, the law must take account of not only the quality of an offender's emotions but also the limited control that individuals have over the shape of their characters. These are points to which we will return in Part IV.

Now consider the requirement that an offender's acts be "voluntary." The purpose of this requirement is to confine punishment to actors who are morally accountable for the harms that criminal law addresses.<sup>318</sup> The law's distinctive conception of voluntariness is in fact much more confounding for the mechanistic view than for the evaluative view of emotions.

---

314. See Robert Nozick, *Philosophical Explanations* 382–84 (1981).

315. See, e.g., Nicola Lacey, *State Punishment* 65–68 (1988) (sketching view of criminal law as system for measuring quality of a person's settled dispositions); Brandt, *supra* note 136, at 169–71.

316. See *supra* text accompanying notes 56–60 (discussing Aristotelian position); notes 150–160 and accompanying text (illustrating evaluative focus on character in voluntary manslaughter doctrine); notes 244–248 and accompanying text (describing murder gradations).

317. See *supra* text accompanying notes 123–127.

318. See Michael S. Moore, *Act and Crime* 46–47 (1993) [hereinafter Moore, *Act and Crime*]; Jeffrie G. Murphy, *Involuntary Acts and Criminal Liability*, in *Retribution, Justice and Therapy* 116, 125–26 (1979) (characterizing punishment of involuntary acts as contrary to justice and morality).

The voluntarist explanation for the voluntary act requirement is straightforward: criminal liability presupposes a voluntary act, for an individual can properly be held responsible for violating the law only if she freely chooses to do so.<sup>319</sup> The problem for this account is that the law views as "voluntary" many acts that are not freely chosen. Acts are considered "involuntary" only if they are attributable to external physical compulsion (such as a shove) or to muscular contractions unmediated by cognition (such as a reflex or an epileptic seizure).<sup>320</sup> However, as the gradations of intentional homicides illustrate,<sup>321</sup> actions attributable to unwilled passions or impulses remain "voluntary" for purposes of the voluntary act requirement.<sup>322</sup> This conception of "voluntariness" is too thin to capture the moral significance that the voluntarist account attributes to choice.<sup>323</sup>

Narrow consequentialism does only slightly better in explaining the voluntary act requirement. For the most part, consequentialism supports confining liability to individuals who voluntarily commit acts that frustrate preferred states of affairs, since deterrence presupposes that individuals have the power to respond to punitive incentives.<sup>324</sup> But this commitment to voluntariness is a qualified one. Punishing individuals even for unwilled conduct might induce them to avoid situations in which they might cause harm involuntarily; in addition, punishing even involuntary behavior avoids the incentive that individuals might have to feign lack of control.<sup>325</sup> Given these competing considerations, it is conceivable that the law's distinctive conception of "voluntary" is optimal from a consequentialist perspective: whereas punishing physically compelled and unconscious acts might be wasteful, punishing impulsive behavior might not be.

The problem with this account is that it is wholly speculative. Whether the benefits of punishing various forms of involuntary conduct exceed the costs of doing so is an empirical matter. Because no empirical support exists, the consequentialist account of the voluntary act requirement presupposes rather than establishes the descriptive power of that theory.<sup>326</sup>

---

319. See, e.g., P.J. Fitzgerald, *Voluntary and Involuntary Acts*, in *Oxford Essays in Jurisprudence* 18-19 (A.G. Guest ed., 1961); Douglas N. Husak, *Philosophy of Criminal Law* 98 (1987); Kadish, *supra* note 129, at 264-67; Murphy, *supra* note 318, at 125-26.

320. See, e.g., Model Penal Code § 2.01(2) (1962); LaFare & Scott, *supra* note 145, § 3.2(c).

321. See *supra* Part I.B-C.

322. See, e.g., *Bratty v. Attorney-General*, [1961] 3 All E.R. 523, 532-33.

323. See Stephen J. Morse, *Culpability and Control*, 142 U. Pa. L. Rev. 1587, 1588-1605 (1994).

324. See, e.g., Model Penal Code § 2.01(1), cmt. x (1985) ("[T]he law cannot hope to deter involuntary movement or to stimulate action that cannot physically be performed . . .").

325. See Hart, *supra* note 128, at 413-15.

326. Cf. Jon Elster, *Nuts & Bolts for the Social Sciences* 7-8 (1989) (distinguishing "story-telling" from genuine explanation). It is instructive to compare this consequentialist

It is possible to develop a much tighter explanation for the contours of the voluntary act requirement if the law is understood to reflect the evaluative conception of emotions. An act is "voluntary" when it is sufficiently engaged with a person's agency to bear moral assessment. The evaluative view assumes that this condition is satisfied so long as a person's conduct can be comprehensibly explained in terms of her beliefs about, appraisals of, and desires for goods essential to her well-being.<sup>327</sup> This is a relatively modest view of human agency; it certainly does not limit the class of voluntary acts to those that are "freely willed" or "chosen." Indeed, under this account it may not be possible to specify fully the constellation of cognitive states that make an act "voluntary"; as Aristotle recognized, the best that can probably be done is to define "voluntariness" negatively by specifying what states of intentionality defeat any attempt to link the actor's behavior to her cognitive appraisals.<sup>328</sup> Many criminal codes take exactly this approach.<sup>329</sup>

### G. *Insanity*

The insanity defense is typically defined in one of two ways. The traditional *M'Naghten* standard focuses only on cognitive impairments, excusing the defendant when, by virtue of a mental disease, she lacks the capacity to understand the nature or wrongness of her acts.<sup>330</sup> Another version of the defense (sometimes referred to as the "irresistible impulse" test) focuses, in addition, on volitional impairments: the defendant is excused when a mental disease impairs her capacity to control her behavior.<sup>331</sup> Although the volitional-impairment standard was once wide-

---

account of the "voluntary act" doctrine with the conventional economic justification for strict liability in tort. Because the tort defendant is required only to compensate the injured plaintiff for actual damages, there is (in theory, and abstracting from various types of litigation costs) no risk that a strict liability standard will cause the defendant to forego socially desirable activity—that is, activity for which the benefits to the defendant exceed the costs to the plaintiff. See Richard A. Posner, *Economic Analysis of Law* 175 (4th ed. 1992). However, criminal penalties are often quite severe and are not capped by any harm to the victim; in addition, such penalties, when they take the form of imprisonment, are extremely costly for society to impose. As a result, the risks of overdeterrence and wasted punishment are much higher when criminal law punishes accidental or unavoidable conduct than when tort law imposes liability for faultless conduct. Cf. *id.* at 239 (concluding that defendants in a strict liability regime will act notwithstanding risk so long as the expected cost of liability is less than the cost of avoidance or of refraining from action). It is thus far from clear—as a matter of theory or otherwise—what precise conception of "voluntary act" is supported by considerations of efficiency.

327. Cf. Martha C. Nussbaum, *The Fragility of Goodness* ch. 9 (1986) (developing this account in the course of explicating Aristotle's position on voluntary acts).

328. See Aristotle, *Ethics*, *supra* note 13, at 1110a30–1111b5, at 41–44.

329. See, e.g., Model Penal Code § 2.01 (1962).

330. See *M'Naghten's Case*, 8 Eng. Rep. 718 (H.L. 1843).

331. See, e.g., Model Penal Code § 4.01(1) (1962). The Model Penal Code formulation, followed by many jurisdictions, requires only that the offender be deemed to "lack [ ] substantial capacity . . . to conform his conduct to the requirements of law." *Id.* The drafters chose this terminology to soften the degree of volitional impairment

spread, the prevailing trend is actually toward the *M'Naghten* approach.<sup>332</sup> Again, we believe that the mechanistic and evaluative conceptions of emotion helpfully illuminate the presuppositions of these tests, and what is at stake in the choice between them.

The volitional-impairment standard is firmly rooted in the mechanistic conception of emotion. It originated in the work of nineteenth-century psychoanalysts, who posited a form of "moral insanity" in which an individual, although conscious of the criminal nature of his behavior, is unable to resist the emotional urge to engage in it.<sup>333</sup> For courts that accepted voluntarist premises, it followed that persons afflicted with such a condition could not be justly punished:

No one can deny that there must be two constituent elements of legal responsibility in the commission of every crime, and no rule can be just and reasonable which fails to recognize either of them: (1) Capacity of intellectual discrimination; and (2) freedom of will. . . . If therefore . . . the disease of insanity can, in its action on the human brain through a shattered nervous organization, or in any other mode, so affect the mind as to subvert the freedom of the will, and thereby destroy the power of the victim *to choose* between the right and wrong, although he perceive it, . . . is such a one criminally responsible for an act done under the influence of such controlling disease? We clearly think not . . . .<sup>334</sup>

---

contemplated by the judge-made "irresistible impulse" test. See Model Penal Code, § 4.01 cmt. at 157 (Tentative Draft No. 4, 1955). For simplicity, we will refer to the volitional-impairment prong of the insanity defense in these jurisdictions interchangeably as the "irresistible impulse test" and the "volitional-impairment test."

332. See Kadish & Schulhofer, *supra* note 169, at 948, 953-54.

333. See Norman J. Finkel, *Insanity on Trial* 30-33 (1988); Edwin R. Keedy, *Irresistible Impulse as a Defense in the Criminal Law*, 100 U. Pa. L. Rev. 956, 956-57 (1952).

334. *Parsons v. State*, 2 So. 854, 859 (Ala. 1887); accord *State v. Reidell*, 14 A. 550, 552 (Ct. Oyer & Terminer Del. 1888) ("No enlightened criminal system of law punishes [individuals] for acts which, if the will was in a healthy state, would never be done . . ."); *Bradley v. State*, 31 Ind. 492, 507 (1870) (if "the power of self-control [is] lost to disease" there is an absence of "free agency" and hence "no moral and legal responsibility" from crime); see also S. Sheldon Glueck, *Mental Disorder and the Criminal Law: A Study in Medico-Sociological Jurisprudence* 232-33 (1925) (arguing that criminal punishment is unjust when defendant lacked volitional capacity); Keedy, *supra* note 333, at 986-87 (arguing that irresistible impulse should be a defense because it negates the voluntary act requirement). The volitional-impairment test is also sometimes defended on the consequentialist ground that persons who experience such impairment are undeterrable. See, e.g., Model Penal Code, § 4.01 cmt. at 156-57 (Tentative Draft No. 4, 1955); see also Abraham Goldstein, *The Insanity Defense* 67-68 (1967) (describing the voluntarist defense of this doctrine). It is far from clear, however, that consequentialism supports this conception of the insanity defense. Rather than excuse volitionally impaired individuals, deterrence considerations could justify punishing them more severely to counteract the strength of their impulses. As one judge put it, "[t]he law says to men who say they are afflicted with irresistible impulses: 'If you cannot resist an impulse in any other way, we will hang a rope in front of your eyes, and perhaps that will help.'" *King v. Creighton*, 14 C. C. C. 349, 350 (1908). In addition, even assuming that certain volitionally impaired offenders

The evaluative conception of emotion, in contrast, does not excuse impulsive behavior merely on account of its impulsiveness. The moral status of emotions, under this view, is determined by the quality of the values they express, not by their effect on volition. Because an individual is responsible for habituating herself to value the right things, she is properly treated as the moral author of her emotions no matter how little power she has to control them. Reasoning along these lines, many turn-of-the-century courts rejected the irresistible impulse test on the ground that it would "utterly pervert and subvert the moral order of things."<sup>335</sup> The claim that the bare intensity of an actor's impulse should excuse him from crime,

may do well enough when applied to the brute world, where there neither is nor can be such a thing as moral obligation, and where individual impulses are regarded as mere instincts without mental control, but it will not do for the government of man, to whom God has given a reasonable soul, by which if he will all his passions may be controlled. . . . The main object of the Penal Code is to compel men to restrain their evil passions and desires, hence the want of such restraint is rather an aggravation of than an excuse for crime.<sup>336</sup>

Modern decisions that reject the volitional-impairment test echo these themes.<sup>337</sup>

---

cannot be deterred by threatened punishment, it cannot be assumed that punishing them will fail to deter others; indeed, recognizing volitional impairment as a defense undermines deterrence by creating the prospect that punishment can be avoided by feigning such incapacity. See Hart, *supra* note 128, at 19–21. For all of these reasons, it is not surprising that few if any common law authorities defended the irresistible impulse test on consequentialist grounds.

335. *Jacobs v. Commonwealth*, 121 Pa. 586, 592 (1888).

336. *Id.* at 592–93; accord *Fitzpatrick v. Commonwealth*, 81 Ky. 357, 361 (1883) ("It is the duty of men who are not insane [under *M'Naghten*] . . . to control their evil passions and violent tempers or brutal instincts, and if they do not do so, it is their own fault, and their moral and legal responsibility will not be destroyed or avoided by the existence of such passions . . ."); *Schwartz v. State*, 91 N.W. 190, 191 (Neb. 1902) (if offender is sane under *M'Naghten*, "he cannot allege the sway of a turbulent passion as an excuse for his crime"); *Flanagan v. People*, 52 N.Y. 467, 470 (1873) (irresistible impulse test "[i]ndulge[s] . . . evil passions [and] weakens the restraining power of the will"); *State v. Brandon*, 53 N.C. 463, 467–68 (1862) ("There are many appetites and passions which by long indulgence acquire a mastery over men more or less strong. Some persons indeed deem themselves incapable of exerting strength of will sufficient to arrest their rule,—speak of them as irresistible, and impotently continue under their dominion; but the law is far from excusing criminal acts committed under the impulse of such passions."); see also *People v. Kerrigan*, 14 P. 849, 851 (Cal. 1887) ("if great moral depravity should be taken as a test of insanity, then the highest degree or enormity of crime would, by virtue of its own atrocity, furnish the best evidence of insanity on behalf of the one who committed the act").

337. See, e.g., *United States v. Torniero*, 570 F. Supp. 721, 729–30 (D. Conn. 1983) (Cabranes, J.) (rejecting volitional-incapacity doctrine on ground that it reflects "deterministic" conception of behavior under which "the very idea of guilt would be corroded"); *State v. Moore*, 76 P.2d 19, 24–25 (N.M. 1938) ("Just as long as human beings live with other human beings in a state of organized society, irritations will occur to arouse anger, jealous[ ]y, and hatred. Nevertheless, human reason, which is supposed to place us

The evaluative conception of emotion does, of course, assume some form of mental competence. This account treats emotions as judgments of value for which a person is morally accountable. Such a view obviously presupposes that the person experiencing the emotion has the cognitive capacity to appreciate her circumstances and to form appropriate judgments about them; if she doesn't, it can't be said that her emotions reflect any sort of intelligible appraisal. Accordingly, the evaluative conception of emotions should generate an insanity defense in line with the *M'Naghten* standard, since that test describes persons who lack the minimum conditions of moral agency presupposed by the evaluative view. Again, legal authorities, old and new, express this evaluative understanding of *M'Naghten*.<sup>338</sup>

This view of *M'Naghten* treats the minimum conditions of moral accountability under the evaluative conception as sufficient conditions for criminal punishment. But as we have stressed, the evaluative view can be combined with a wide variety of theories about the appropriate dimensions of criminal law.<sup>339</sup> Some of these accounts might indeed excuse mentally incapacitated offenders who wouldn't be treated as insane for purposes of *M'Naghten*. A broader formulation of insanity—even one that incorporates the volitional-impairment test—would remain consistent with the evaluative conception of emotion so long as it did not purport to excuse actors based on the intensity of their emotions alone.

Such consistency can be achieved, for example, through the doctrinal requirement that the defendant's incapacity stem from a mental disease. Courts have traditionally stressed the distinction between mental disease as a "legal" concept and mental disease as a "medical" concept.<sup>340</sup> But they have also traditionally refused to specify exactly what the legal concept of "disease" is, thus delegating to the jury the task of identifying which forms of mental incapacity entitle an offender to be excused.<sup>341</sup> Accordingly, even under the volitional-impairment conception of insanity, the jury is free to limit the defense to offenders whose diseases negate the character-shaping power that the evaluative view assumes, or even to those whom the jury views as possessing virtuous character.

---

in a position superior to other animals, makes us realize that we are not to fly from anger to violence without punishment unless we would destroy the compatible association of man with man.").

338. See, e.g., *Torniero*, 570 F. Supp. at 731 ("The legal condition of insanity, then, occurs when the state of the defendant's mind is such that it is alienated from ordinary human experience. We cannot understand the outlook of the insane person; the barrier of mental disease or defect interrupts the possibility of the jury's comprehension."); *People v. Coleman*, 1 N.Y. Crim. Ct. 1, 2 (1881) (the defendant who satisfies the *M'Naghten* standard "is devoid, both in morals and in law, of the elements essential to the constitution of crime, and hence is an object of pity and protection, and not of punishment").

339. See *supra* p. 305.

340. See, e.g., *State v. Guido*, 191 A.2d 45, 52-53 (N.J. 1963).

341. See *id.*



Empirical assessments suggest that juries do exactly that. Studies consistently find that mock juries are no more likely to acquit under one formulation of insanity than another.<sup>342</sup> However instructed, the jury assesses the defendant's sanity according to a lay construct that focuses on a wide array of extra-doctrinal considerations, including "the defendant's background," his "relationship with the victim," his "intent to harm,"<sup>343</sup> and his "culpab[ility] *before the act* for bringing about [his] incapacity."<sup>344</sup> These factors assure that even the volitional-impairment test of insanity is unlikely to be applied mechanistically.

Indeed, the history of the volitional-impairment test bears out its consistency with the evaluative conception of emotion. For example, the "irresistible impulse" conception of insanity was one of the vehicles (along with self-defense) of the traditional "unwritten rule" that a husband could lawfully execute his wife's paramour (or rapist)<sup>345</sup>—an application of the doctrine portrayed in the 1950s novel and popular film, *Anatomy of a Murder*.<sup>346</sup> "[W]e can perceive," one court of the time coyly remarked,

where a man of good moral character such as that possessed by the defendant, highly respected in his community, having regard for his duties as a husband and the virtue of women, upon learning of the immorality of his wife, might be shocked, or such knowledge might prey upon his mind and cause temporary insanity. In fact it would appear that such would be the most likely consequence of obtaining such information.<sup>347</sup>

Of course, where the husband, too, is an adulterer, the same court explained in another case, the prosecution is free to introduce evidence of *his* "moral delinquency . . . to rebut the inference that the mental shock or strain [of his wife's infidelity] was sufficiently great to unseat his reason."<sup>348</sup> Whatever the voluntarist presuppositions of the courts that fashioned the "irresistible impulse" test, the juries and courts that have applied it have always paid at least as much attention to defendants' characters as to their psyches.

---

342. See, e.g., Norman J. Finkel, *De Facto Departures from Insanity Instructions*, 14 *Law & Hum. Behav.* 105, 112–13 (1990); Norman J. Finkel et al., *Insanity Defenses: From the Jurors' Perspective*, 9 *Law & Psychol. Rev.* 77, 83–84 (1985); James R.P. Ogloff, *A Comparison of Insanity Defense Standards on Juror Decision Making*, 15 *Law & Hum. Behav.* 509, 521, 524 (1991).

343. Ogloff, *supra* note 342, at 526.

344. Norman J. Finkel & Sharon F. Handel, *How Jurors Construe "Insanity"*, 13 *Law & Hum. Behav.* 41, 57 (1989).

345. See, e.g., Note, *The Unwritten Law as Defense to Homicide*, 19 *Neb. L. Bull.* 146, 148–49 (1940); Weinstein, *supra* note 277, at 229.

346. See Robert Traver, *Anatomy of a Murder* (1958).

347. *Hamilton v. State*, 244 P.2d 328, 335 (Okla. Crim. App. 1952); see also *Abbott v. Commonwealth*, 55 S.W. 196, 198 (Ky. 1900) (defendant understandably driven insane by victim's seduction and impregnation of defendant's sister and victim's subsequent attempt to abandon the sister following forced marriage).

348. *Ditmore v. State*, 293 P. 581, 583 (Okla. Crim. App. 1930).

This transmutation of the irresistible impulse test generalizes. Even when the law insists that they express themselves in mechanistic terms, decisionmakers are likely to think and judge in evaluative ones. Concepts like "impulse," "irresponsibility," "involuntary," and "premeditation" are extraordinarily vague. They furnish ready cover for implicit emotional evaluations; indeed, it is often unclear how these concepts could possibly do any work without such hidden appraisals. Frequently, then, what turns on the choice between these mechanistic terms and their overtly evaluative counterparts—"passion," "character," "unreasonable," "evil"—is not what the law will be in substance, but only what it will be in form. Or, to put it another way, the issue is not whether the law will be evaluative or mechanistic, but only whether it will be honest or dishonest.

### III. SHIFTING ASSESSMENTS OF EMOTION IN CRIMINAL LAW

Keith Peacock returned home unexpectedly one evening and discovered his wife making love with another man. He shot her dead several hours (and several shots of whiskey) later. Upon Peacock's plea of guilty to voluntary manslaughter, Judge Robert Cahill sentenced Peacock to eighteen months' imprisonment in a work release program. The judge expressed sympathy for Peacock, stating that he could imagine nothing that would provoke "an uncontrollable rage greater than this: for someone who is happily married to be betrayed in your personal life, when you're out working to support the spouse."<sup>349</sup> "I seriously wonder how many men married five, four years," Cahill continued, "would have the strength to walk away without inflicting some corporal punishment."<sup>350</sup>

The only remarkable thing about this story is that anyone viewed it as remarkable. For centuries, adultery has been regarded as a provocation adequate to mitigate the killing of either the paramour or the defendant's unfaithful wife to voluntary manslaughter; for centuries men convicted of this offense have been sentenced leniently.<sup>351</sup> Nevertheless, public reports of Judge Cahill's sentence and remarks ignited a firestorm of controversy. Newspapers across the country ran critical editorials.<sup>352</sup>

---

349. Sheridan Lyons, Court Panel to Probe Judge in Sentencing, *Baltimore Sun*, Oct. 20, 1994, at 1B.

350. *She Strays, He Shoots*, Judge Winks, *N.Y. Times*, Oct. 22, 1994, at A22 [hereinafter *She Strays*].

351. See *supra* text accompanying notes 345–348. See generally Donna K. Coker, *Heat of Passion and Wife Killing: Men Who Batter/Men Who Kill*, 2 *S. Cal. Rev. L. & Women's Stud.* 71 (1992) (surveying historical use of voluntary manslaughter in cases in which cuckold kills unfaithful wife).

352. See, e.g., *A Judge Who Dishonors the Bench*, *Hartford Courant*, Oct. 21, 1994, at A16; *Justice Mocked in Maryland*, *Boston Herald*, Oct. 24, 1994, at 26; *'Passion Killing' Sentence Absurd*, *Sun-Sentinel* (Fort Lauderdale, Fla.), Oct. 22, 1994, at 18A; *She Strays*, *supra* note 350, at A22; Ann G. Sjoerdsma, *18 Months for a Wife's Life*, *Chi. Trib.*, Nov. 14, 1994, at 21; *Unequal Justice*, *St. Louis Dispatch*, Nov. 1, 1994, at 12B.

Protesters picketed the courthouse, calling for Cahill's removal.<sup>353</sup> Members of the Maryland General Assembly introduced a resolution condemning him.<sup>354</sup> Formal disciplinary proceedings were initiated.<sup>355</sup> And the Maryland judiciary agreed to withhold new sentencing guidelines so that provisions relating to domestic violence could be reviewed and possibly strengthened.<sup>356</sup>

What was once settled in the law—that marital infidelity can provoke a reasonable man to homicidal rage—is now contested. But why? Has the public come to doubt the genuineness or intensity of the cuckold's anger? Or does a substantial portion of the public now question the moral quality of this emotion, and hence the appropriateness of treating the cuckold's anger as reasonable? We believe the story of Keith Peacock and Judge Cahill illustrates another descriptive benefit associated with the evaluative conception of emotion—namely, its power to explain the responsiveness of legal assessments of emotion to changes in social norms.

The evaluative conception (at least as we use it) assumes that an individual's emotional life is in large part socially constructed.<sup>357</sup> Many of the valuations embodied in a person's emotions are imparted to her by others at an early age. Equally important, individuals experience emotions against a rich background of social norms that define what goods are appropriately valued and by whom. A man who resents a woman's interest in another man, for example, experiences understandable *jealousy* if she is his wife or lover, but inappropriate *envy* if she is a person with whom he has no established relationship.<sup>358</sup>

Social norms are especially central to anger.<sup>359</sup> A person experiences anger when she perceives that another has slighted her in a significant way. This perception presupposes conventions that specify from whom a person may legitimately demand respect (from her social subordinates; from her peers; from certain members of her family; from all members of the community) and what forms of behavior count as disrespectful (insulting words; the failure to include the person in some important activity; an inappropriate sexual overture). For this reason, anger can be viewed as a mechanism by which a person defends her status in

---

353. See John W. Frece, Ouster of Judge Sought, *Baltimore Sun*, Dec. 8, 1994, at 1B; Sheridan Lyons & Robert G. Matthews, Oust Judge Cahill, Protesters Urge, *Baltimore Sun*, Oct. 22, 1994, at 1B.

354. See Editorial, *Baltimore Sun*, Dec. 13, 1994, at 18A.

355. See Editorial, *supra* note 354, at 18A; Frece, *supra* note 353, at 1B; Lyons, *supra* note 349, at 1B.

356. See Janet Naylor, Maryland Judges Delay Relaxation of Sentence Guidelines, *Wash. Times*, Jan. 26, 1995, at A1.

357. See *supra* text accompanying notes 106–115.

358. See Rom Harré, An Outline of the Social Constructionist Viewpoint, in *The Social Construction of Emotions*, *supra* note 107, at 2, 6–7.

359. See Averill, *supra* note 65, at 55–72.

the community and the social norms on which her status depends.<sup>360</sup> By acknowledging or failing to acknowledge a person's anger (or lack of anger) as appropriate, onlookers, too, reinforce particular norms and destabilize others.<sup>361</sup>

The norms against which anger and other emotions are constructed vary widely across cultures and can change radically over time within a single culture.<sup>362</sup> Indeed, the transformation of social norms is a conspicuous part of contemporary political and social life. This is especially so in the domains of gender and sexuality, where traditional, hierarchical norms are now highly contested and in some cases completely discredited.<sup>363</sup> Even if marital infidelity continues to be viewed as a serious moral wrong, most persons cringe at the suggestion that it is "the highest invasion of [a man's] property."<sup>364</sup> And because most are likely to understand the violent anger of Keith Peacock as reflecting exactly that assessment, it is unsurprising that Judge Cahill's sentence and remarks provoke so much more hostility today than they would have in the eighteenth century.<sup>365</sup>

The public expects the law to track these shifting social norms,<sup>366</sup> and it frequently does. If the common law paradigm of appropriate rage was the cuckold, the emerging but highly contested contemporary paradigm is the battered woman, or more generally, the person who is victimized or abused by someone who owes that person a special duty of concern and respect.<sup>367</sup>

360. See Theodore D. Kemper, *Power, Status, and Emotions: A Sociological Contribution to a Psychophysiological Domain*, in *Approaches to Emotions* 369, 371-74 (Klaus R. Scherer & Paul Ekman eds., 1984).

361. Cf. Elster, *supra* note 178, at 99-100 (noting role of shared emotional responses in sustaining social norms).

362. See, e.g., Rom Harré & Robert Finlay-Jones, *Emotion Talk Across Times*, in *The Social Construction of Emotions*, *supra* note 107, at 220; Heelas, *supra* note 107, at 234-36, 258.

363. See generally William H. Chafe, *The Paradox of Change: American Women in the 20th Century* ch. 10 (1991) (noting change in public attitudes and opinions toward gender roles since 1945); Peter G. Filene, *Him/Her/Self: Sex Roles in Modern America* (2d ed. 1986) (providing historical account of evolving definitions of gender roles since the mid-nineteenth century); Sheila M. Rothman, *Woman's Proper Place: A History of Changing Ideals and Practices, 1870 to the Present* (1978) (describing different conceptions of woman's proper place and the impact of those conceptions on social policies during the period from 1870 to 1978).

364. *Regina v. Mawgridge*, 84 Eng. Rep. 1107, 1115 (1707).

365. Consider, for example, the text of the proposed legislative resolution condemning Judge Cahill: "We have long discarded those ancient and intolerable laws, customs and practices that allowed a husband to use force and violence against a wife for indiscretions, real or suspected; and . . . [w]e should not return to those dark ages of gross indecencies against women." Editorial, *supra* note 354, at 18A.

366. For a general account of how changes in norms relating to sexuality have affected public expectations about the grade of different criminal offenses, see Robinson & Darley, *supra* note 177, at 160-69.

367. As another judge remarked (understatedly) in the aftermath of the Peacock sentencing, "[t]hese kinds of things—cases that have to do with sexual offenses or physical

In some instances, this evaluative shift has been accommodated and promoted by formal changes in doctrine. Until 1973, the Texan who killed his wife's paramour had a complete statutory defense to homicide.<sup>368</sup> As late as the 1950s, citizens of the state understood the "paramour statute"—which was regularly invoked—to embody the legitimate right of a cuckold to vindicate his honor through violence.<sup>369</sup> But by the time it was repealed, the statute was regarded as "an anachronism—a frontier idea whose time had gone" and which rendered "the state a legal laughing stock."<sup>370</sup> In 1979, the Texas Court of Criminal Appeals ruled that the belief that deadly force was necessary in self-defense must be assessed from the perspective of "an ordinary and prudent man" and not from the point of view of a reasonable battered woman.<sup>371</sup> That decision was overturned in 1991 when Governor Ann Richards signed legislation (vetoed by her predecessor) establishing the right to introduce expert testimony relating to the defendant's state of mind in any case in which a victim of domestic abuse (whether an adult or a child) is charged with killing her abuser.<sup>372</sup>

More often, the law's responsiveness to shifting social norms has been mediated by the discretionary judgments of courts and juries, which predictably fit evolving mores into the spaces that the law creates for evaluative assessments. Ironically, the vehicle for such assessments is at least sometimes criminal law doctrines that appear mechanistic. Told to acquit the defendant on grounds of insanity only if he experienced an "irresistible impulse," or on grounds of self-defense only if he was impelled to use deadly force by "primal impulse" of "self-preservation," the jury has historically applied these defenses to the person who it believes has behaved virtuously, albeit lawlessly.<sup>373</sup> And the identity of the virtuous out-

---

abuse of a spouse—are something that strikes a chord with the public today." Lyons, *supra* note 349, at 1B.

368. See Texas Penal Code art. 1220 (repealed 1973).

369. See John Bainbridge, *The Super-Americans: A Picture of Life in the United States, As Brought into Focus, Bigger than Life, in the Land of Millionaires—Texas* 238–39, 248–249 (1961).

370. Correspondence from Albert Alschuler to Dan M. Kahan (May 5, 1995) (on file with the Columbia Law Review). Alschuler was one of the official reporters for the committee that proposed repeal of the paramour statute as part of the comprehensive reform of the Texas Penal Code in 1973.

371. See *Valentine v. State*, 587 S.W.2d 399, 401 (Tex. Crim. App. 1979). But cf. *Fielder v. State*, 756 S.W.2d 309, 321 (Tex. Crim. App. 1988) (suggesting that admission of expert testimony on previously abused defendant's state of mind is sometimes relevant).

372. See Tex. Code Crim. Proc. § 38.36(b)(2) (West Supp. 1996). See generally Meredith J. Duncan, Comment, *Battered Women Who Kill Their Abusers and a New Texas Law*, 29 Hous. L. Rev. 963, 966–67 (1992).

373. See *supra* text accompanying notes 345–348.

law is changing. He is no longer Major Frederick Manion;<sup>374</sup> she is now Lorena Bobbitt.<sup>375</sup>

#### IV. DEFENDING THE EVALUATIVE CONCEPTION OF EMOTION IN CRIMINAL LAW

We believe that the evaluative conception is true and the mechanistic conception false as accounts of what emotions are. If we are right, then the truth of the evaluative conception is one reason for the criminal law to prefer it. The criticisms made of the mechanistic conception show it to be not simply inadequate, but actually internally confused and incoherent. It stands to reason that legal rules built upon the mechanistic view are likely to be confused and incoherent as well.

Nevertheless, we recognize that the philosophical truth of the evaluative conception does not conclusively establish the superiority of this view in criminal law. Doctrines that embody the evaluative conception require decisionmakers to appraise the valuations embodied in the emotions of criminal offenders and victims. The law need not make such appraisals—at least not overtly. If there were some reason to exclude this type of assessment from law, criminal doctrines could give emotions the significance attributed to them by the mechanistic conception despite the falsity of that view as a philosophical matter.

The evaluative conception of emotion in criminal law thus requires an independent normative defense. Supplying one is the aim of this Part. We begin by showing that doctrines structured to reflect the evaluative conception better promote the recognized purposes of punishment than do doctrines structured to reflect the mechanistic view. We then anticipate important objections to embedding moral appraisals of emotions in law. Finally, we offer an important qualification of the evaluative conception. The moral acceptability of condemning an individual offender for the content of her emotions, we believe, presupposes some mechanism for assessing that person's responsibility for her character. This assessment can be justly and effectively achieved, without undermining legal evaluations of offenders' emotions, through institutionalizing *mercy* in criminal sentencing.

##### A. *The Evaluative Conception and the Purposes of Criminal Law*

The idea that a single normative theory does or should determine the shape of all criminal doctrines is exceedingly implausible. Because

---

374. See Traver, *supra* note 346.

375. See David Margolick, *Lorena Bobbitt Acquitted In Mutilation of Husband*, N.Y. Times, Jan. 22, 1994, at 1. It has become increasingly common for juries to acquit victims of domestic violence who kill their abusers when the formal requirements of the insanity or self-defense doctrines are not met. See Maria L. Marcus, *Conjugal Violence: The Law of Force and the Force of Law*, 69 Cal. L. Rev. 1657, 1724–25 (1981); Rosen, *supra* note 285, at 410–11; Donald L. Creach, Note, *Partially Determined Imperfect Self-Defense: The Battered Wife Kills and Tells Why*, 34 Stan. L. Rev. 615, 629, 631 (1982).

the subject matters regulated by criminal law are extraordinarily diverse, theories that generate compelling prescriptions in one setting are almost certain to generate unacceptable ones in others. This is not to say that normative theorizing in criminal law is pointless. It is only to say that successful theorizing is likely to consist of eclectic and contestable syntheses of different justifications, any one of which accounts only imperfectly for our considered moral judgments.<sup>376</sup>

The ultimate plurality of theories in criminal law provides a hospitable environment for the evaluative conception of emotion. The evaluative conception is not itself a theory of punishment; it requires only that the law, however justified, take account of the moral quality of emotions. And this prescription is consistent with almost all important theories. Indeed, no matter which theories are used, and in no matter what proportions they are combined, the evaluative conception is likely to prove superior to the mechanistic conception in advancing the basic aims of criminal law. To illustrate this contention, we will address three commonly asserted justifications of criminal punishment: expressive condemnation; deterrence; and individual desert.

1. *Expressive Condemnation.* — The expressive theory of punishment can be viewed as an instance of a more general expressive account of human behavior and institutions.<sup>377</sup> This account stresses that actions have *meanings* as well as *consequences*.<sup>378</sup> Social norms enable rational behavior by defining how persons (or communities) who value particular goods—whether the welfare of other persons, their own honor or dignity, or the beauty of the natural environment—should behave. Actions that conform to, or defy, these norms thus express a person's (or a community's) attitude toward these goods.

This theory supports a distinctive conception of wrongdoing. To be wrongful, an act must do more than adversely affect someone's tangible interests. A competitor's marketing of a superior product, for example, can harm a merchant financially as much as the theft of her goods. The reason that theft but not competition is wrongful is that theft expresses—in a way that competition (ordinarily) does not—a false assessment of the

---

376. See generally Henry M. Hart, Jr., *The Aims of the Criminal Law*, 23 *Law & Contemp. Probs.* 401 (1958) [hereinafter *Hart, Aims*]; Robinson, *Hybrid Principles*, *supra* note 132, at 19, 41–42 (proposing the construction of a formula for balancing conflicting purposes of criminal sanctions).

377. For general treatments of the expressive theory, see Anderson, *supra* note 191; Cass R. Sunstein, *Incommensurability and Valuation in Law*, 92 *Mich. L. Rev.* 779, 820–24 (1994). For expressive theories of punishment, see Joel Feinberg, *The Expressive Function of Punishment*, in *Doing and Deserving: Essays in the Theory of Responsibility* 95, 95–118 (1970); Jeffrie G. Murphy & Jean Hampton, *Forgiveness and Mercy* (1988) [hereinafter *Murphy & Hampton, Forgiveness*]; Nozick, *supra* note 314, at 370–80; Hart, *Aims*, *supra* note 376; Dan M. Kahan, *What Do Alternative Sanctions Mean?*, 63 *U. Chi. L. Rev.* (forthcoming 1996).

378. See Lawrence Lessig, *The Regulation of Social Meaning*, 62 *U. Chi. L. Rev.* 943 (1995).

merchant's worth. Against the background of social norms, the thief's behavior communicates to the merchant and to others that she views the merchant's interests as unworthy of her respect.<sup>379</sup>

Punishment also has a unique signification. To punish, a community must do more than impose suffering or deprivation. A person can lose as much liberty, for example, through military conscription as she can through imprisonment. The reason that imprisonment but not military service counts as punishment is that imprisonment conveys society's authoritative moral condemnation.<sup>380</sup> By imposing the appropriate form and degree of affliction on a wrongdoer, the political community reaffirms its commitment to the values that the wrongdoer's own act denies.<sup>381</sup>

Because criminal law expresses condemnation, what a political community punishes, and how severely, tell a story about whose interests are valued and how much.<sup>382</sup> That such significance is often attached to the law can be seen in the recurring complaint that lenient treatment of certain offenses—whether domestic violence or hate crimes—shows that the well-being of certain persons just “doesn’t count” in the eyes of the law.<sup>383</sup> The expressive theory thus reveals how much is at stake politically and morally in the correspondence between the message that a wrongdoer's act conveys and the law's punitive retort.

If the law is to perform this expressive function, then doctrines should be structured to reflect the evaluative rather than the mechanistic conception of emotion. To determine what a person's actions *mean*, it is necessary to consider not just the consequences of her behavior but also her reasons, including her emotional motivations, for engaging in such acts.

---

379. See Jeffrie G. Murphy, *Forgiveness and Resentment*, in Murphy & Hampton, *Forgiveness*, *supra* note 377, at 14, 25; Jean Hampton, *Forgiveness, Resentment and Hatred*, in Murphy & Hampton, *Forgiveness*, *supra* note 377, at 35, 43–44.

380. See Hart, *Aims*, *supra* note 376, at 404–06.

381. See Jean Hampton, *The Retributive Idea*, in Murphy & Hampton, *Forgiveness*, *supra* note 377, at 130.

382. See *id.* at 140–42.

383. See, e.g., Scott Armstrong, *Case Against Simpson Intensifies Death-Penalty Debate in US*, *Christian Sci. Monitor*, Sept. 6, 1994, at 2 (reporting comment by feminist activist that prosecutor's decision on whether to seek the death penalty for O.J. Simpson “raises the issue of whether a battered woman's life is as important as a celebrated man's life”); Judge Draws Protests After Cutting Sentence of Gay Man's Killer, *N.Y. Times*, Aug. 17, 1994, at A15 (reporting comment by gay activist that light sentence of man convicted of manslaughter for intentional killing of homosexual “s[ays] . . . that it is O.K. to kill faggots”); Lyons & Mathews, *supra* note 353, at 1B (reporting reaction of female protester that 18-month, work-release sentence of man who killed wife after discovering her infidelity sends message that “murder is no big deal—in fact, is a fitting punishment” for unfaithful wives). For an extended account of the stake that victims and those who identify with them have in criminal trials, see George P. Fletcher, *With Justice for Some: Victims' Rights in Criminal Trials* (1995).



The expressive upshot of emotions is clearest in homicide cases. To return to an earlier example, the reason that a mother who kills the sexual abuser of her daughter is less worthy of condemnation than the man who kills a person whom he believes to be a homosexual is that her emotional motivation expresses less reprehensible valuations than does his: her anger appropriately values the worth of at least something in her situation—namely, her daughter's well-being—whereas his hatred appropriately values nothing. Under an expressive theory, the man should be punished more severely to repudiate the more reprehensible message of homophobia implicit in his act. A mechanistic formulation of voluntary manslaughter that looked only to the intensity and not to the quality of offenders' emotions would sever the correspondence between the valuations expressed in wrongdoers' actions and the condemning retort of punishment.

The same point can be made about the role of emotions in defenses. It is conceivable that the rage of Bernhard Goetz and the fear of a battered woman can be reduced to psychological urges of equivalent intensity. But their emotional motivations for using deadly force express appraisals of their circumstances that it would be morally inappropriate to equate through a mechanistic formulation of self-defense. Likewise, a woman who acquiesces in the abuse of her children to avoid a threatened physical attack upon herself may be prompted by a fear every bit as intense as that of a woman who agrees to commit an armed robbery to avoid a threat to her or her family. However, a mechanistic formulation of duress that excused both would ignore morally relevant distinctions in what their respective fears express about how much each woman values her own welfare relative to that of loved ones and that of strangers.<sup>384</sup>

Even when mechanistic doctrines generate the right results, moreover, they send the wrong message. The rationale of such doctrines is that intense emotions impair rational agency, making it difficult or impossible for actors to choose morally preferred outcomes. This is a profoundly misleading account, however, of why the angry mother and the second fearful woman discussed in our preceding examples are entitled to exculpation or mitigation. If they are, it is not because their emotions have destroyed their capacity to behave like rational moral agents, but precisely because their emotions express rational and morally appropriate assessments of their situations. Thus, structuring the doctrine to reflect the mechanistic conception of emotion not only mutes the expressive upshot of punishment; it systematically falsifies it.

2. *Optimal Deterrence.* — Optimal deterrence seeks to maximize the return on society's investment in punishment. From this point of view, it makes sense for society to expend punishment resources if, but only if, the resulting social gain from averted criminality exceeds the cost of pun-

---

384. See *supra* text accompanying notes 300–301.

ishment.<sup>385</sup> In addition, society should allocate its punishment resources across different forms of criminality so as to maximize the net amount of social harm averted.<sup>386</sup> Optimal deterrence also suggests that the law should be structured to encourage those emotions that promote socially desired outcomes and to discourage those that frustrate them.<sup>387</sup>

It should be clear that optimal deterrence presupposes some form of consequentialism. In Part II, we criticized as descriptively inadequate a narrow form of consequentialism that treats social wealth (or any other independently specified state of affairs) as normative for the law's assessment of emotions. Our disagreement, however, is not with consequentialism or optimal deterrence *per se*. Indeed, we now want to defend a deterrence theory that is grounded in a richer form of consequentialism, one that values emotions in proportion to the intrinsic worth of the valuations they express. Doctrines structured along these evaluative lines are more likely to generate optimal regulation of behavior motivated by emotions; in addition, such doctrines are better suited to inculcating socially desirable emotional dispositions.

a. *Optimal Regulation of Emotional Behavior.* — When combined with the mechanistic conception of emotion, a voluntarist account of criminal doctrines clearly leads to underdeterrence. The voluntarist account equates moral responsibility with choice. Accordingly, it counsels mitigation of punishment in proportion to the strength of the offender's emotional motivations. Deterrence theory suggests exactly the opposite: all else equal, the stronger an actor's impulse or desire to engage in a forbidden act, the more severe the punishment must be in order to discourage her from acting.<sup>388</sup> There is no similar tension between evaluative doctrines and this insight of deterrence theory, for the evaluative view does not treat the intensity of a person's emotion apart from its moral quality as a reason for mitigating punishment.

Whereas the voluntarist account risks insufficient punishment of behavior motivated by undesirable emotions, the narrow consequentialist account risks excessive punishment of conduct motivated in whole or in part by desirable ones. This is so because the consequentialist account, when informed by the mechanistic conception of emotion, lacks an intelligible and morally acceptable theory of value.

Optimal deterrence presupposes some understanding of which outcomes are socially valued and disvalued, and how much. Without such a theory of value, it would be impossible to identify which kinds of conduct

---

385. See, e.g., Bentham, *supra* note 133, at 162, 169–71; Michael & Wechsler, *supra* note 137, at 1269.

386. See Bentham, *supra* note 133, at 171.

387. See *supra* text accompanying note 137.

388. See Posner, *An Economic Theory of the Criminal Law*, *supra* note 133, at 1223; see also Bentham, *supra* note 133, at 170 (equating "strength of the temptation" with the "profit of the offence" and concluding that severity of punishment "must therefore rise with the strength of the temptation").

should be deterred. It would also be impossible to make sensible decisions about how to allocate limited punishment resources among different forms of wrongdoing.

Any consequentialist theory of value that is specified independently of the evaluative view of emotion is likely to generate morally unacceptable outcomes. Consider, for example, the proposition that society should punish battered women more severely than other killers precisely because they are so strongly impelled to kill by rage or fear. This claim is unacceptable if we believe that the battered woman's emotions appropriately evaluate the threat that the batterer poses to her well-being. In that case, her acts should not be deterred at all. Or recall once more the cases of the man who kills a homosexual and the mother who kills her daughter's abuser. If (as we suggest) the mother's anger appropriately values something in her situation, while the man's hatred appropriately values nothing in his, then it would be unacceptable to punish them equally simply because their emotional motivations are equally intense; persons moved to kill by homophobic hatred cause greater social harm, and thus warrant (all else equal) a greater portion of society's limited punishment resources, than do mothers moved to kill by love for their children.

In sum, to be morally acceptable consequentialism must take appropriate account of the intrinsic worth of emotional valuations. Accordingly, the evaluative conception is an essential supplement to the deterrence theory because without it it is impossible to determine what states of affairs the law should be trying to maximize.<sup>389</sup>

This argument assumes that evaluative assessments are properly viewed as the basis, and not the product, of any consequentialist theory of value.<sup>390</sup> We believe they are. Certainly, such evaluations are revisable; the assessments we make of emotions change as the values underlying them evolve, as our empirical knowledge grows, and as we critically re-examine particular assessments in light of others. Nevertheless, our emotional evaluations are too basic to our apprehension of the moral world to be subordinated to an independent theory of value that purports to be normative for all of them all at once.<sup>391</sup>

b. *Formation of Appropriate Emotional Dispositions.* — A prospective offender's propensity to commit a crime is a function, roughly and abstractly, of (1) her perception of the cost of committing it and (2) the strength of her desire to do so. Accordingly, deterrence can be achieved

---

389. Cf. Sen, Rights and Agency, *supra* note 134, at 29–30 (arguing that consequentialism is morally acceptable only when it incorporates certain agent-relative valuations into desired states of affairs).

390. Cf. Anderson, *supra* note 191, at 29–30 (states of affairs have value only by virtue of conformity to independent expressive norms); Sunstein, *supra* note 377, at 821 (arguing that consequentialist theories of value are derived from expressive judgments arrived at independently of consequentialism).

391. See Nussbaum, Love's Knowledge, *supra* note 50, ch. 6. See generally Elster, *supra* note 178, at 125–51 (arguing that normative assessments grounded in social norms cannot be explained in terms of their contribution to individual or collective utility).

by either (1) raising the expected cost of crimes through threatened punishment or (2) reducing prospective offenders' desires to commit them.<sup>392</sup>

Criminal law has traditionally been conceived of as doing both. The latter is often referred to as the "moralizing" or "educative" effect of punishment.<sup>393</sup> The basic mechanism behind it is preference adaptation.<sup>394</sup> Because punishment expresses authoritative moral condemnation, it stigmatizes desires to engage in criminal behavior as deviant. To avoid the discomfort or dissonance associated with holding such desires, individuals internalize dispositions, outlooks, and tastes that conform to the social norms expressed in criminal prohibitions.<sup>395</sup> Individuals thus refrain from criminality not because they fear the threatened punishment but because they have no desire to engage in such behavior; and they have no desire to engage in such behavior because they know it is deemed worthy of criminal punishment.<sup>396</sup>

The conviction that it is immoral to disobey the law makes a substantial contribution to deterrence.<sup>397</sup> Individuals who desire to engage in criminal acts can be deterred only by their perception of the severity of

392. See Kenneth G. Dau-Schmidt, *An Economic Analysis of the Criminal Law as Preference-Shaping Policy*, 1990 *Duke L.J.* 1, 22-23; Jack P. Gibbs, *Preventive Effects of Capital Punishment Other than Deterrence*, 14 *Crim. L. Bull.* 34, 40-41 (1978).

393. See, e.g., Johs Andenaes, *General Prevention—Illusion or Reality?*, 43 *J. Crim. L., Criminology, & Police Sci.* 176, 179 (1952); Jean Hampton, *The Moral Education Theory of Punishment*, 13 *Phil. & Pub. Aff.* 208, 212 (1984); Gordon Hawkins, *Punishment and Deterrence: The Educative, Moralizing, and Habitutive Effects*, 1969 *Wisc. L. Rev.* 550, 550. For empirical assessments, see Leonard Berkowitz & Nigel Walker, *Laws and Moral Judgments*, 30 *Sociometry* 410, 417-21 (1967); Nigel Walker & Michael Argyle, *Does the Law Affect Moral Judgments?*, 4 *Brit. J. Criminology* 570, 575 (1963-1964); Nigel Walker & Catherine Marsh, *Do Sentences Affect Public Disapproval?*, 24 *Brit. J. Criminology* 27, 38-39 (1984).

394. See Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* ch. 3 (1983); Cass R. Sunstein, *Legal Interference with Private Preferences*, 53 *U. Chi. L. Rev.* 1129, 1146-50 (1986).

395. See Johannes Andenaes, *The General Preventive Effects of Punishment*, 114 *U. Pa. L. Rev.* 949, 978 (1966); Elliot Aronson & J. Merrill Carlsmith, *Effect of the Severity of Threat on the Devaluation of Forbidden Behavior*, 66 *J. Abnormal & Soc. Psychol.* 584, 584-85 (1963).

396. Consider Sir James Fitzjames Stephen's account of the moralizing effect of capital punishment:

Some men, probably, abstain from murder because they fear that if they committed murder they would be hanged. Hundreds of thousands abstain from murder because they regard it with horror. One great reason why they regard murder with horror is that murderers are hanged with the hearty approbation of all reasonable men.

Sir James Fitzjames Stephen, *A General View of the Criminal Law of England* 99 (1863).

397. See Tom R. Tyler, *Why People Obey the Law* chs. 4-5 (1990); Harold G. Grasmick & Donald E. Green, *Legal Punishment, Social Disapproval and Internalization as Inhibitors of Illegal Behavior*, 71 *J. Crim. L. & Criminology* 325, 328 (1980); Paul R. Robinson & John M. Darley, *Utility of Desert* (unpublished manuscript, on file with authors).

legal sanctions and the probability of their imposition—information that is very imperfectly disseminated.<sup>398</sup> In addition, the stronger the desire of a given population to engage in criminal acts, the more society must invest in punishment in order to deter; were the desire to engage in murder, rape, or theft widespread and intense, the cost of administering effective criminal sanctions would be intolerably high.<sup>399</sup> Accordingly, it is essential that there be reliable social mechanisms for inculcating aversions to such behavior.

If criminal law is expected to be one such mechanism, then it must send intelligible signals about what kinds of tastes and outlooks—including emotional dispositions—a well-constituted person should have. It does so when structured evaluatively. In that event, the law expressly communicates a story about character—the characters of those who lash out violently, and the characters of those against whom they use violence.

Mechanistically structured doctrines transmit much more ambiguous signals. Such doctrines are distinguished by their disavowal of any assessment of the moral quality of emotions. Consequently, even when mechanistic doctrines convey what an individual should or should not do, they fail to convey what kind of person she should be. Indeed, the persons whom the mechanistic theory purports to excuse (at least when combined with a voluntarist account) are not the ones with good characters, but the ones without characters. They are persons who have lost the ability to act like responsible agents. Their acts tell us nothing about how good persons view the world.

3. *Individual Desert.* — The theory of individual desert counsels punishment in strict proportion to fault. This position is thoroughly anti-consequentialist. Faulty individuals must be punished even when doing so produces no discernible future benefits to society, and nonfaulty ones must not be punished even when doing so would advance important collective ends such as social wealth or expressive condemnation.<sup>400</sup>

How do emotions affect assessments of fault? The mechanistic and evaluative conceptions support radically different answers to this question. At least when combined with the voluntarist theory of moral assessment, the mechanistic view suggests that judgments of fault must be abstracted from emotions. What matters is *choice*, and since emotions are unwilling they figure in moral assessment only to the extent that they dis-

---

398. See generally Gibbs, *supra* note 392, at 35 (noting lack of empirical evidence to support assumption that members of public accurately perceive probability of punishment); Raaj K. Sah, *Social Osmosis and Patterns of Crime*, 99 J. Pol. Econ. 1272, 1273 (1991) (discussing studies showing that members of the public have widely divergent perceptions of the probability of punishment for criminal offenses).

399. See Tyler, *supra* note 397, at 22–23; Dau-Schmidt, *supra* note 392, at 22.

400. See, e.g., Michael S. Moore, *The Moral Worth of Retribution*, in *Responsibility, Character, and the Emotions*, *supra* note 147, at 179, 182 (stating that under the retributive theory of punishment, “the institutions of punishment . . . are justified by the rightness or fairness of the institution in question, not by the good consequences such institution may generate”).

place choice, in which case a person should be relieved in part or in whole from the assignment of praise or blame.<sup>401</sup> The evaluative conception, in contrast, treats individuals as the moral authors of the appraisals embodied in their emotions; a person is accountable not only for making good choices but for having good character, which consists in experiencing appropriate rather than inappropriate passions.

The evaluative view's focus on character better accounts for common moral intuitions than does the mechanistic view's narrow focus on choice.<sup>402</sup> We typically do credit individuals, morally, for the positive qualities of their emotional motivations. A person moved to care for a sick and elderly relative out of love, for example, merits greater praise than a person who provides the same care out of hope of financial reward. This assessment, moreover, is unaffected by the conclusion that the emotional underpinnings of love are deep-seated and unwilling; the person who passively, even instinctively experiences these sensibilities deserves admiration. Likewise, certain emotions conventionally provoke negative moral assessments. We condemn, for example, persons who experience racial hatred. And again, it seems just to hold a person accountable for such sensibilities even when he experiences them without choice.

These intuitions and practices may in part reflect the Aristotelian premise that individuals ultimately choose their characters. As the Pennsylvania Supreme Court put it in explaining why the common law concerns itself with the quality as well as the intensity of a person's passions, "[m]an is largely the creature of education; his character depends principally, if not wholly, upon his own will, and for that character he is legally responsible."<sup>403</sup> The racist, for example, might be unable to suppress his hate by an act of will on any particular occasion. But insofar as adults have the power and opportunity to reform the values that their emotions embody, the racist's sustained failure to revise his makes him an appropriate object of condemnation.

These considerations suggest that evaluative doctrines are more likely to generate results consistent with notions of individual desert than are doctrines grounded in the mechanistic conception of emotion. Nevertheless, we believe that the justifications on which such intuitions rely are incomplete; they tend to understate the social contribution to, and hence overstate the plasticity of, individual emotional dispositions. The lingering tension between the evaluative view and the moral importance of choice is an issue to which we will return in discussing the role of mercy in criminal sentencing.

---

401. See *supra* text accompanying notes 128-129.

402. For extended discussions of the relationship between emotional dispositions and moral appraisal, see Lawrence A. Blum, *Friendship Altruism and Morality* 169-207 (1980); Justin Oakley, *Morality and the Emotions* 38-85, 160-87 (1992).

403. *Small v. Commonwealth*, 91 Pa. 304, 308 (1879).

## B. *Moral Dilemmas*

Doctrines that reflect the evaluative conception of emotion are judgmental. They attach significance not just to the consequences of offenders' actions, but to the quality of the values embodied in their emotional motivations. When the homophobe is convicted of murder rather than voluntary manslaughter, the severity of punishment condemns the disrespect of his victim's identity expressed in his hatred; when the law permits a man to repel an attack with deadly force and ignore a safe route of retreat, the doctrine of self-defense endorses the conception of honor embodied in his aversion to shame.

In this section, we will address two specific objections relating to the morally judgmental quality of the evaluative view: that the evaluative view is illiberal; and that it risks entrenching unjust moral evaluations.

1. *Is the Evaluative Conception Illiberal?* — The mechanistic view seems to steer clear of value judgments. In speaking about the quality of an offender's emotion, it asks only about strength and relation to volition; such notions seem conveniently neutral. By contrast, the evaluative view is, obviously, evaluative: it says that some things are worth getting very upset about and others are not; it appraises the evaluations internal to the offender's emotions as reasonable or not reasonable. In the process it takes a whole range of moral stands—about the importance of the murder of a child, about the (alleged) difference between the killing of an unfaithful wife's lover and the killing of an unfaithful girlfriend's lover, about the appropriateness or inappropriateness of the disgust occasioned by witnessing a homosexual act.

Such stand-taking may trouble us not only because we worry that the *wrong* stands are likely to be taken (a point that we address presently), but because we think that it is improper for the law to get involved in such judgments *at all* in a liberal democratic society. Surely people should be free to live their lives by their own conceptions of the good, deciding how much importance to accord to children, how much respect to the moral or religious standards that give rise to disgust at homosexual conduct, and so forth. Surely the law should be studiously neutral among competing conceptions of the good—and this seems to be a strong reason to prefer the thin sort of assessment of emotion promoted by the mechanistic view.

It is unnecessary for us to take a position on the general question of how far and in what ways a liberal regime should indeed be neutral about the good. We focus only on the question at issue, which is how these questions should be confronted in the context of the criminal law. In other words, we do not say anything one way or another about how a regime such as ours should deal with questions about family values, the morality of homosexual conduct, and so forth. We ask only how such questions of value should be handled when they arise in our dealings with people who, by general consensus, have committed a criminal act. The question is whether at that point we need to ask, or ought to ask, about

the quality of the evaluations exhibited in the emotion with which the act is done.

The answer, we believe, must be yes, because it simply could not be otherwise. The mechanistic view does not really avoid the question of evaluation; it just handles it in a clumsy and offhand way. By saying that the only distinctions in emotion that are going to count are distinctions of strength or intensity, the view does take a stand on the good, albeit a perverse one. It holds, without arguing, that getting angry over the murder of one's own child and getting angry over the sight of two same-sex strangers making love differ in no salient way, that the only thing worth attending to here is the strength of the offender's anger, which might, of course, happen to be equally great in the two cases. That's not neutrality—that's a decision to equate what our intuitions usually do not equate, to disregard what deeply rooted norms of reasonableness make central.

This disregard, furthermore, has social consequences. When the mechanistic approach is combined with a voluntarist approach, for example, the wrong people will be treated with indulgence, and the message of deterrence sent to the community as a whole, to potential offenders in particular, will be a confused and dangerous one. The message will be: don't worry about making yourself into the sort of person who gets provoked only by events that are really grave, don't worry about being the "reasonable person," don't worry about schooling yourself not to lash out in unreasonable ways—just make sure that if you do, you can show that you had really strong feelings about it. It's all right to hate your neighbor for some morally irrelevant reason and to be provoked to crime by this hatred—just make sure that when you commit your crime you really have intense passions, and then the law will treat you lightly. This is not a neutral message; it is a message that fosters and gives comfort to racism and homophobia and other reprehensible feelings.

The mechanistic view of emotions is no more successful in extricating the law from contentious evaluations when it is combined with consequentialism. There is no morally neutral theory of value; even a pure deterrence theory of punishment still has to decide what deserves to be punished and how much. It is possible to imagine a regime that gives the same punishment to the homophobic killer and the parent who kills the abuser of her child based on their equal disregard for the lives of their victims abstracted from all else, including the differing valuations expressed in their respective emotions. But this regime would itself rest on a radical theory—indeed a radical theory of the good—since both deeply shared intuitions and good moral arguments support a distinction in these cases.

But even if the mechanistic conception of emotion could somehow reduce the judgmental quality of criminal law, that would not make it more congenial to the best understanding of what liberalism requires. For the evaluative view is much more strongly conducive to the ideal of



public reason, and thus provides crucial support for a deliberative liberal regime.<sup>404</sup>

Operating with the mechanistic view requires us to treat large numbers of our fellow citizens as beyond the parameters of rational judgment, and, indeed, to regard the whole business of shaping one's character and emotions as a matter of happenstance, rather than as part of a rational public culture. This is certainly not a view of persons that lends support to liberal constitutionalism of the sort that Rawls and other liberal theorists wish to defend.<sup>405</sup>

Under the evaluative view, by contrast, we are continually (except in a small number of cases of extreme mental disorder) treating our fellow citizens as reasonable, and promoting in society at large a view that human beings are or should be reasonable and rational. This is both more appropriate as a view of one's fellow citizens under liberal constitutionalism and more effective in promoting the stability of such a regime.<sup>406</sup> Moreover, such a view opens to public debate and deliberation many crucial areas of human conduct—what provocations are reasonable, what anger a reasonable citizen would experience—and thus fosters the construction and maintenance of a deliberative public culture. The mechanistic view of persons, by permitting us to ask only how strong certain impulses are, would actively discourage public deliberation on such matters.

Finally, were we really to try to do away with value judgments in the law, it would mean a radical and more or less unimaginable change. If we are not to be permitted to say that the murder of one's own child is a more appropriate provocation than the sight of two strangers making love, on what basis are we to be permitted to classify crimes and to judge that some are graver than others? When we think about crimes such as murder and rape, but also fraud, larceny, blackmail—we are making value judgments all the time, classifying different levels of offense by their gravity, and recommending different levels of sentence for different offenders within a given offense. In fact, without evaluation it would be difficult to know how we would even describe offenses: the terms "murder" and "rape" and "blackmail" are far from neutral about the good, and if we would assiduously avoid them we would have to find a new language for the description of these acts—in terms, perhaps, of the sheer bodily movements performed. But nobody has even tried to show how this

---

404. See John Rawls, *Political Liberalism* ch. 6 (1993).

405. See *id.* at 18–20 ("Beginning with the ancient world, the concept of the person has been understood, in both philosophy and law, as the concept of someone who can take part in, or who can play a role in, social life, and hence exercise and respect its various rights and duties."); see also John Rawls, *A Theory of Justice* 504–10 (1971) (the "capacity for moral personality" is defined by both the capacity to have a conception of the good and the capacity to have a sense of justice).

406. See Rawls, *supra* note 404, at 240–54.

might be done, or why we should want to do it.<sup>407</sup> It is inconsistent and unfair of the liberal neutralist to object to our use of appraisal in connection with the evaluative conception, and yet to allow it to remain a ubiquitous feature of our system of criminal (and, of course, also civil) law.

We believe, then, that no reasonable liberalism can be neutral about the good to the extent and in the ways that would be promoted by the dominance of the mechanistic view (which, as we have said, is in any case only pseudo-neutral). And no regime of law remotely like ours could survive such pseudo-neutrality.

2. *The Problem of "Bad Morality" and a (Partial) Institutional Solution.* — Another potential objection is that the evaluative conception risks reinforcing "bad morality." Even if it is appropriate in theory for the law to examine the moral quality of offenders' emotions, it cannot be assumed that decisionmakers will invariably make appropriate assessments in practice. What would prevent a homophobic court or jury from concluding that homosexual behavior is adequate provocation to mitigate murder to manslaughter? Indeed, regarding the law as evaluative in nature might be thought to compound the injustice of such a result precisely because it treats the jury's verdict as an endorsement of a defendant's emotional motivation. The way to avoid entrenching bad moral valuations, it might be argued, is to structure doctrines mechanistically rather than evaluatively.

This objection is also unpersuasive. The short answer is that recognizing the evaluative conception of emotion in criminal law cannot really make the problem of bad morality worse. The longer and slightly more contentious answer is that it can actually make it better, particularly if responsibility for assessing emotional motivations is properly allocated among different decisionmakers.

a. *No Worse, Possibly Better.* — The evaluative conception cannot make the problem of bad morality worse because even doctrines that are overtly mechanistic furnish ready vehicles for implicit assessments of the quality of offenders' emotional motivations. The "irresistible impulse" conception of insanity is mechanistic, yet historically the jury has used it to excuse defendants, such as the cuckold, whose emotional motivations it approves.<sup>408</sup> The classic formulation of self-defense is also mechanistic, purporting to identify the circumstances in which the use of deadly force is impelled by the "primal impulse" of "self-preservation."<sup>409</sup> Yet consider the acquittal of Bernhard Goetz, who all but admitted that he shot his African-American victims to vindicate his honor as the victim of past mug-

---

407. This difficulty is shown clearly in the Stoic view of punishment; Stoic moral theory holds that all damages to persons are trivial and of no worth; it therefore becomes a difficult matter to understand how offenses will be categorized, or why murder and rape should even be punished at all. See Nussbaum, *Therapy*, *supra* note 15, chs. 11–12.

408. See *supra* text accompanying notes 345–348.

409. See *supra* text accompanying notes 263–264.

gings.<sup>410</sup> Similar appraisals are made, yet concealed, when juries apply the seemingly mechanistic doctrine of "premeditation."<sup>411</sup>

In sum, in political communities in which bad morality predominates, mechanistic doctrines will not stop the jury from crediting inappropriate emotional motivations. They only drive those assessments underground. But do such assessments at least do less harm there than they do on the surface of the law?

The answer, we believe, is no. The import of mechanistic decisions is rarely lost on those who are affected by them most directly. The jury's endorsement of Goetz's sense of honor, for example, did not elude African-Americans;<sup>412</sup> nor did women in Maryland fail to appreciate the anachronistic conception of female virtue underlying Judge Cahill's lenient sentence of Keith Peacock.<sup>413</sup> Under such circumstances, the only parties from whom the law's evaluative content is possibly concealed are the decisionmakers themselves, who are spared the need to confront their appraisals directly, and less interested members of the public, who lack the information and incentive to pierce through the veneer of mechanistic rhetoric.

The evaluative view eliminates this selective distortion, and in so doing mitigates the problem of bad morality. For two reasons, legal evaluations of offenders' emotions are likely to deviate least from the moral ideal when those evaluations are most fully exposed to view.

First, the evaluative view forces decisionmakers to accept responsibility for their moral assessments and to give reasons for them in a public way. Before a homophobic killer could prevail on a manslaughter theory, for example, both the judge and the jury would have to *endorse* his hatred as "reasonable."<sup>414</sup> Even assuming they privately share his biases, they might be unwilling to take the public act of aligning themselves with him if that is the only available interpretation of their decision.<sup>415</sup> The mechanistic view, in contrast, purports to be concerned only with the intensity and not the quality of an offender's emotional motivation. Because it thus enables decisionmakers publicly to disavow any endorsement of the defendant's emotional motivations, the mechanistic view furnishes decisionmakers with an excuse to indulge prejudices that they might have felt

---

410. See *supra* text accompanying notes 277–278.

411. See *supra* text accompanying notes 245–258.

412. See generally Fletcher, *Self-Defense*, *supra* note 278, ch. 11 (analyzing reactions of African Americans to Goetz verdict).

413. See *supra* text accompanying notes 349–356.

414. See *supra* text accompanying notes 150–174.

415. For an economic model of this phenomenon, see Timur Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification* chs. 2–3 (1995). Kuran argues that individuals participating in collective decisionmaking face reputational incentives to suppress preferences that they perceive (correctly or incorrectly) to be contrary to shared ideals or norms. See *id.*; see also Frederick Schauer, *Giving Reasons*, 47 *Stan. L. Rev.* 633, 657–58 (1995) (arguing that decisionmakers are less likely to rely on bias or self-interest when required to give public account of the reasons for their decisions).

constrained to reject had their decision been framed as one of moral principle.

Second, acknowledging the evaluative underpinnings of the law fully exposes decisionmakers' assessments to the public. This outcome promotes critical examination, and possibly even the repudiation, of moral appraisals that the public is unwilling to accept. The mechanistic view is less likely to promote such deliberation precisely because the message it sends about the quality of the offenders' motivation is muted.

To illustrate this point, consider a case that is in fact the mirror image of our homophobic killer example. In 1988, a state judge in Texas imposed a relatively light sentence on a man convicted of murdering two homosexuals who had allegedly solicited that man.<sup>416</sup> The judge explained his decision this way: "I put prostitutes and gays at about the same level, and I'd be hard put to give somebody life for killing a prostitute."<sup>417</sup> His decision was outrageous; and it provoked public outrage. The judge was formally censured, and ultimately defeated at the polls.<sup>418</sup> In the wake of this and other incidents, moreover, the legislature enacted a hate crimes statute that expressly enhances the penalty for crimes motivated by bias against any group.<sup>419</sup>

The outcome in this case would have been just as egregious—and no doubt just as demeaning to gays—had the judge based his sentence on the ground that the defendant had suffered an overwhelming "impulse" to kill. But it seems unlikely that a decision based on that mechanistic ground would have goaded the public as effectively as the patently evaluative (and hence patently offensive) one actually offered by this judge.

b. *A (Partial) Institutional Solution.* — We have suggested that evaluative doctrines have the potential to counteract the problem of bad morality by exposing the resolution of contentious issues to plain view. This conclusion has an institutional upshot: in the distribution of the authority to make assessments of emotional valuations, there ought to be a preference for decisionmakers whose judgments are most fully amenable to public scrutiny.

The legislature best fits this description. The legislative process is more accessible to public participation than is the judicial process and typically generates much clearer resolutions of contentious issues than do individual trials. Admittedly, it is impossible to specify in advance, by stat-

---

416. Lisa Belkin, *Texas Judge Eases Sentence for Killer of 2 Homosexuals*, N.Y. Times, Dec. 17, 1988, at 8.

417. *Id.*

418. See *Gay Rights Groups Hail Defeat of Judge in Texas*, N.Y. Times, Dec. 4, 1992, at B20; *Judge is Censured Over Remark on Homosexuals*, N.Y. Times, Nov. 29, 1989, at A28.

419. See *Tex. Penal Code Ann.* § 12.47 (West 1994); *Tex. Code Crim. Proc. Ann.* art. 42.014 (West Supp. 1996); see also Clay Robison, *Richards Signs Hate Crimes Bill into Law*, *Houston Chronicle*, June 20, 1993, State section, at 3 (noting that purpose of the legislation is to enhance "criminal offenses motivated by the victims' race, religion, ethnicity, sexual orientation or national origin").

ute, the effect to be given to all emotional motivations for engaging in criminal acts; the types of valuations that emotions embody, and the assessments that we make of them, are too numerous, too diverse, and too fact-specific to be comprehensively addressed by general rules. But it is perfectly feasible for a legislature to specify by statute the effect to be given to at least *some* identifiable and recurring emotional motivations.<sup>420</sup> Killings motivated by homophobia and other forms of group animus, for example, ought (in our view) to be included expressly in the definition of first degree murder.

To the extent that evaluations must be based on the facts of particular cases, courts should assume an aggressive role. In particular, they should screen out claims that rest on manifestly inappropriate emotional valuations; they should not submit such claims to juries, because juries are the decisionmakers whose judgments are least subject to public scrutiny. Thus, we approve of the limited use of categorical definitions of adequate provocation for purposes of voluntary manslaughter. In the absence of legislative action, courts should take the lead in declaring that homosexual advances and like conduct are inadequate provocation as a matter of law rather than permitting juries to decide this issue as a matter of fact.<sup>421</sup> Likewise, they should declare that the infidelity of a man's wife is no longer legally sufficient to mitigate murder (of either the wife or the paramour) to manslaughter, given the traditional and continuing nexus between anger in these circumstances and hierarchical conceptions of gender.<sup>422</sup>

The preference for evaluation by judges rather than juries also affects other doctrines, including self-defense. Courts should assess the quality of a defendant's emotional motivations, for example, in considering which types of expert testimony are relevant to the "reasonableness" of a defendant's perceptions. Recognizing battered woman syndrome does not require recognizing "battered subway commuter syndrome" or other asserted conditions that rest on inappropriately low valuations of the lives or behavior of nonculpable victims.<sup>423</sup> Likewise, courts should be more aggressive in precluding self-defense claims in cases, such as *Goetz*, in which evidence that is insufficient as a formal matter is nonetheless likely to move the jury to credit inappropriate emotional evaluations.

Our claim here is not that judges are wiser or more just than juries. As our Texas example illustrates, judges may be just as likely as juries to make bad evaluations. But as this same example also suggests, when judges do express inappropriate evaluations, citizens often take notice, and they often take action. Jury verdicts, which are much harder to interpret, provoke this kind of reaction much more rarely.

---

420. See, e.g., Pillsbury, *Evil and Murder*, *supra* note 248, at 480.

421. See *supra* note 172 and accompanying text (discussing state of law on this issue).

422. See Horder, *supra* note 147, at 192-94.

423. See Dershowitz, *supra* note 8, at 18-19 (listing "abuse excuses," some valid, some invalid).

C. *Conviction and Sentencing: A Role for Mercy?*

Finally, we arrive at what is in many respects the deepest and most interesting challenge to our view. It is the claim that the evaluative conception is too sanguine about people's possibilities, and, in consequence, unreasonably harsh. For it demands of people not only that they conform their conduct to a certain standard, but also that they shape their characters, and the quality of their emotions, in accordance with prevailing norms of reasonableness. It treats the anger of a criminal offender as something for which that person is appropriately appraised. Carr's disgust at the sight of lesbian lovemaking is judged inappropriate, and Carr is treated by the law as someone who should have formed his character differently. The woman who lashes out against the abuser of her child, by contrast, is treated as someone who has a reasonable character—and, once again, the assumption is made that this character is hers, that she is appropriately assessed for the quality of evaluation intrinsic to her emotion.

So far in the paper we have flinched before these implications, saying that we will look only at what behavior manifests, and not raise the metaphysical question of freedom. But can we really avoid raising this question when we observe that people form their emotions in circumstances not of their own making, some of which may be quite deforming?

Richard Wright's novel *Native Son* vividly poses this question.<sup>424</sup> The novel depicts the life of Bigger Thomas, an impoverished and uneducated African-American who eventually commits two violent criminal acts, at least one of which (the killing of his lover Bessie) is clearly a murder. Throughout the novel, Wright prevents us from having easy sympathy for Bigger by showing him as someone whose emotions are deformed and inappropriate. His actions are dominated by overwhelming rage, shame, and fear, and he cannot be relied upon to behave either legally or morally. He is indeed a dangerous killer, and his crimes express blameworthy attitudes toward his victims and his surroundings. He is clearly guilty of murder, at least in the one case, and there is no reason to see his emotions as reasonable responses to an adequate provocation. At the same time, Wright makes us experience discomfort with our urge to condemn Bigger by showing us in detail how his emotions have been shaped by both poverty and racism, how shame at the color of his skin, fear of the dominant white community, and rage at his unequal and immobilized situation all interact in the daily events of Bigger's life. We are led to think that this is a person who did not have the degree of control over his character development that we usually do, given the extremely closed and unequal situation in which he lives—that his basic human potential has been stunted and deformed by his adverse situation.

---

424. Richard Wright, *Native Son* (1940). To find a real-life example of the same dilemma, however, we would not be forced to look beyond the case reporters. See *United States v. Alexander*, 471 F.2d 923, 957-58 (D.C. Cir. 1973).

In short, there are a lot of reasons why people develop unreasonable emotions. Isn't it too harsh to assume that an individual is always responsible for his character? We believe that the answer is yes; Bigger Thomas did not create his situation—we all did—and we cannot now justly treat him as if he were the only source of the evil that his actions embody. But can we afford this concession? Once we recognize that someone like Bigger cannot be blamed for his bad character, aren't we back to some form of voluntarism in which the right question to ask is simply whether the forces that bore down on the person were sufficient to overwhelm choice?

Our answer to this challenge is, and should be, complex.<sup>425</sup> Put succinctly, it is that the objection is obtuse. Moral and legal assessments are complex, and operate at different levels. Neither consists of one simple question, "Is this person responsible or not?"; rather, both involve a fine-tuned series of questions that help us to distinguish assessments that are appropriately made of *a person's actions* (including its emotional motivations) and *a person's responsibility for her character*.

Begin with nonlegal moral practices. When someone disregards a friend's interests in a serious way, the friend is likely to be angry and to condemn her. The offended party won't necessarily excuse the transgression upon learning that the source of this inattention was an emotion (say, an intense desire to succeed in work that caused the offending party to ignore the friend for long stretches). Indeed, the emotional grounding of such behavior might compound or even constitute the offense, since it reveals that the offending party values some good essential to her well-being more than she ought to value her friend's interests. But if the offending party is indeed a friend, the aggrieved party is unlikely to end his appraisal with a bare assessment of the quality of her actions. He will judge his friend's failing against the background of what he knows about her entire character; and he will likely listen to her account of why she failed him, including her emotional motivations. If the account moves him, then he is likely to forgive the transgression and continue the friendship, particularly if the offending party herself acknowledges that she has wronged him.

To generalize, moral assessment frequently involves two steps. In the first, we appraise the quality of a person's actions, including its emotional motivations. At that point, we are largely unconcerned with issues of responsibility for character; we are concerned only to appraise the act for what it is—either good or bad, beautiful or ugly. But at the second stage, we are intensely interested in the offending party's responsibility; we are open to a narrative account of why she fell below the moral norm, and we may in fact make an additional, distinct appraisal of her based on that

---

425. See Martha C. Nussbaum, *Equity and Mercy*, 22 Phil. & Pub. Aff. 83 (1993) [hereinafter Nussbaum, *Equity*].

story.<sup>426</sup> These two stages of assessment are frequently marked by social norms or conventions that serve to separate them and to prevent them from undermining each other. To *forgive* a friend's inattention, for example, is not to take back the initial *condemnation* of her behavior; on the contrary, the convention of forgiveness presupposes blame, for if the offending conduct were no longer recognized as wrong, there would be nothing to forgive.<sup>427</sup>

This two-stage structure—in which we frequently both blame a person severely for a wrong done, and then, looking at things in another light, come to see with sympathy the road that led this person to become a person of the sort—is ubiquitous. We strongly encourage it in moral education, in which children are both taught moral values, and, often at least, invited to love their enemies—to try to exercise charity of mind and sympathetic imagination toward those who have wronged them. The same two stages are deeply situated in the history of our political culture. Lincoln's Second Inaugural, for example—with its scathing excoriation of the practice of slavery, combined with its appeals to mercy and charity<sup>428</sup>—offers us a classic example of a way in which the two-stage process has figured in the construction of a regime that is both moral and humane. Lincoln shows us that there is nothing incoherent about combining “firmness in the right” with a merciful attitude toward those who have committed a grievous wrong; he argues that only such an attitude will enable us to maintain a stable public culture.

The same two stages often inform legal assessment through the formal distinction between conviction and sentencing. In determining an offender's guilt or innocence, and the grade of her offense, the law evaluates her actions, including her emotional motivations; and at that point, the law—at least if it takes an evaluative stance on emotions—is ordinarily unconcerned with how the defendant came to be the way she is. But during the sentencing process, the law *has* traditionally permitted the story of the defendant's character-formation to come before the judge or jury in all its narrative complexity, in such a way as to manifest any factors hidden in the background of this life that might, once presented, give rise to sympathetic assessment and to a merciful mitigation of punishment.<sup>429</sup> A long tradition of merciful sentencing in the criminal law—

---

426. Cf. John Sabini & Maury Silver, *Emotions, Responsibility, and Character*, in *Responsibility, Character, and the Emotions*, supra note 147, at 165, 172–73 (noting that judgments of character have a dual dimension, one grounded in responsibility and morality and one grounded in aesthetics and emotion).

427. See Murphy, supra note 379, at 20–25.

428. See Abraham Lincoln, Second Inaugural Address (Mar. 4, 1865), in 1 *Documents of American History* 442, 442–43 (Henry S. Commager ed., 9th ed. 1973) (“Both read the same Bible and pray to the same God . . . With malice toward none, with charity for all, with firmness in the right . . . let us . . . bind up this nation's wounds.”).

429. See, e.g., *Woodson v. North Carolina*, 428 U.S. 280, 303–05 (1976) (noting that “the prevailing practice of individualizing sentencing determinations generally reflects simply enlightened policy”).



beginning, in fact, from the Roman Stoics<sup>430</sup>—has connected mercy with sentencing, rather than with the determination of guilt, and has carved the process of adjudication into these two phases.<sup>431</sup> In fact, the very concept of mercy implies this two-stage assessment: for, unlike compassion or sympathy, mercy (like forgiveness) is a kind of leniency that presupposes fault. A merciful judge is one who, for whatever reasons, sets the level of penalty more leniently than the nature of the offense, taken strictly, demands.<sup>432</sup>

What does the mercy tradition have in mind in making this division of assessment into two phases? It would appear that the rationale is something like what we have already suggested. For reasons of determining guilt or innocence, it is right to look into the quality of the agent's motivations and emotions at the time of the offense, but not to raise large cumbersome issues about the person's whole history and process of character-formation. One may justify this division on pragmatic grounds such as efficiency, imperfect knowledge,<sup>433</sup> and deterrence;<sup>434</sup> one may justify it as the procedure most likely to be fair to defendants from a wide range of backgrounds;<sup>435</sup> or one may justify it on the expressive ground that the bad and ugly should be recognized as such,<sup>436</sup> regardless of how the offender came to be implicated in them.

On the other hand, since we know that human life is not so simple and that people encounter obstacles of many kinds on the way to forming their characters, it is right that we should have, *and express*, a certain anxiety about this situation. So there is also a place for the narrative history—especially when the penalty that may be fixed is a severe one, and especially where the defendant's background shows some *prima facie* evidence of unusual hardship or inequality. In such a case, we want to look into things more deeply, to see whether we may have missed some unusual impediment that deformed the process of character formation. At this point, a long moral and legal tradition holds that we owe it to the dignity and humanity of the defendant to let the entire history appear, in case some aspect might inspire a merciful response:

A process that accords no significance to relevant facets of the character and record of the individual offender or the circumstances of the particular offense excludes from consideration in fixing the ultimate punishment of death the possibility of com-

---

430. See, e.g., Seneca, *On Anger*, supra note 15, at 1.19.5–6, at 38; Seneca, *On Mercy*, supra note 81, at 1.20.2, at 152.

431. See Seneca, *On Mercy*, supra note 81, at 1.20–22, at 152–54.

432. See Nussbaum, *Equity*, supra note 425, at 85–87, 92–105 (discussing Aristotle's and Seneca's views of justice, equity, and mercy); Murphy & Hampton, *Forgiveness*, supra note 377, at 158–60.

433. See, e.g., Seneca, *On Anger*, supra note 15, at 2.29.2–4, at 67.

434. See, e.g., *id.* at 1.19.5–6, at 38; Seneca, *On Mercy*, supra note 81, at 1.22.3, at 154.

435. See, e.g., Seneca, *On Mercy*, supra note 81, at 2.7.2–5, at 164.

436. See, e.g., *id.* at 1.2.1–2, at 130–31.

passionate or mitigating factors stemming from the diverse frailties of humankind. It treats all persons convicted of a designated offense not as uniquely individual human beings, but as members of a faceless, undifferentiated mass to be subjected to the blind infliction of the penalty of death.<sup>437</sup>

This is the spirit in which we would try to resolve the dilemma posed by Wright in *Native Son*. Bigger Thomas had committed horrible crimes, the reprehensibility of which consisted at least in part of the false valuations expressed in his emotions. Failing to punish him, severely, would show insufficient commitment to the values that his actions denied. But to take into account only what his acts (including his emotional motivations) expressed would be to say something false about his situation as well. Bigger is not entirely at fault; many factors share the blame for the malformation of his character. By showing leniency in his punishment, we acknowledge the importance of those factors and affirm our society's duty to do better by persons in his situation.

Allowing this space for mercy in Bigger's case does not retreat to a voluntarist position. It is neither necessary nor sufficient to find that Bigger's circumstances "caused" his behavior in some mechanistic fashion. The reason that mercy is appropriate is that lenience supplements and enriches the disposition of his particular case. In another, mercy might confer no comparable benefits and might indeed impoverish the statement made by conviction, even assuming that the offender's unfortunate upbringing made an essential contribution to his crime. Against the background of dissensus and conflict surrounding sexual orientation, for example, there might be no room for mercy in Carr's case.

Indeed, it seems that mercy is rooted much more firmly in expressive considerations than in voluntarist ones. In a criminal case as complex and ambiguous as Bigger Thomas's, we expect the law to make a wide range of statements. We want it to condemn the crime, including the reprehensible emotional motivations underlying it. We want it to reaffirm the worth of the victim, whose value has been falsely denied. But we also demand (or should demand) that the law acknowledge "the diverse frailties of human kind." This requires, according to Seneca and the philosophical mercy tradition, taking up a certain attitude toward ourselves—saying, perhaps, we are all weak and subject to deformation, and had we

---

437. *Woodson v. North Carolina*, 428 U.S. 280, 304 (1976) (opinion of Stewart, Powell, and Stevens, J.J.). See also the defense of this use of narrative in the penalty phase in *Walton v. Arizona*, 497 U.S. 639 (1990), and the related discussion of sympathy in the penalty phase in *California v. Brown*, 479 U.S. 538 (1987), both in the majority opinion by Chief Justice Rehnquist (who cites *Woodson* approvingly), *id.* at 541, and in the more detailed discussion of the tradition in Justice Brennan's dissent, *id.* at 548–51 (Brennan, J., dissenting). In that case, both the majority and the dissenters agreed about the appropriateness of appeals to mercy at the penalty phase; they disagreed only about whether the California jury instruction would mislead jurors concerning the types of attitudes they might and might not exercise. See Martha C. Nussbaum, *Poetic Justice: The Literary Imagination and Public Life* 77–78 (1995).

been in that person's shoes, who knows what we would have done; it requires, too, an outcome that avoids separating ourselves totally from those who have wronged us, and that emphasizes instead the values of imagination and mutual aid.<sup>438</sup> A disposition that purports to answer only a single, abstract question—did the defendant's background "cause" his crime? or even does the defendant "deserve" to be punished?—will never be rich enough to convey all of these meanings. One that divides act assessments and character-formation assessments at least comes closer.

It is obvious, nevertheless, that the line between these two stages of assessment is to some extent oversharped and arbitrary. If character-formation gives grounds for mercy, doesn't it also make us see the time of the crime in a new light? Yes, and no. No, in the sense that we still want to insist that such defendants are appropriately assessed for what they have done. No matter what their childhood was like, they are still different from insane people, and we want to mark that difference—for reasons of deterrence, fairness, and expressive condemnation. On the other hand, yes, in the sense that we see the act itself with a certain sympathy as having grown out of an unusually deformed or deprived formative process; and it is this sympathy that is manifested in the merciful waiving of the harshest punishment in the sentencing phase.

We recognize, in addition, that this two-stage approach does not perfectly describe existing legal practices. In most (perhaps even all) jurisdictions, act evaluations and character-responsibility evaluations are not rigidly cordoned off from each other and assigned, respectively, to the guilt/innocence and sentencing phases. In particular, judges have conventionally been free to make evaluative assessments of offenders' actions (including their emotional motivations) at sentencing.<sup>439</sup> Indeed, nothing prevents assessments made at this point from colliding with and completely undermining assessments made in the adjudication of guilt. Consider, for example, the expressive import of the Texas judge's lenient sentencing of the man convicted of killing in homophobic hatred.

Accordingly, the perfect implementation of this two-stage approach may require modest reform of existing practices. Just as conventions are needed in ordinary moral life to separate action evaluations from character-formation evaluations, so such conventions are needed in law to separate evaluative condemnation at the guilt phase from mercy in sentencing, and to prevent them from undermining each other. The best means to accomplish this objective, we believe, is a regime of qualified determinate sentencing: the factors relevant to sentencing should be specified as fully as possible in advance, subject to discretion to mitigate punishment based on circumstances that detract from an offender's responsibility for

---

438. See Nussbaum, *Equity*, *supra* note 425, at 99–105.

439. See, e.g., *McCleskey v. Kemp*, 481 U.S. 279, 308 (1987) (endorsing "discretion 'on . . . the particularized characteristics of the individual defendant' " (quoting *Gregg v. Georgia*, 428 U.S. 153, 206 (1976) (plurality opinion of Stewart, Powell, and Stevens, JJ.)).

her character; the discretion must be strictly limited, however, to avoid unduly lenient sentences that undermine the judgment of condemnation expressed in the conviction. This is currently how at least some courts understand the Federal Sentencing Guidelines.<sup>440</sup>

Determining a person's responsibility for her own emotional life confronts us with a difficult problem, to which it would be simplistic to suppose there is a neat or perfect solution. The two-stage process of assessment we propose is admittedly imperfect. It schematically divides into two discrete questions what is really a complex network of questions about character and state of mind. It seems to us, however, that any good theory in this area must respect the ultimate tension that pervades our moral beliefs about responsibility and character; any that even purported to dispel this tension would inevitably misunderstand it. Precisely because our proposal recognizes its own incompleteness, the doctrine of mercy in sentencing does justice to the complexity of the ways in which agency and constraint, choice and necessity, are interwoven in a human life.<sup>441</sup>

#### CONCLUSION

We began this Article by noting the criminal law's apparently confused and ambivalent responses toward emotions. Our goal was to test this appearance—to determine how much sense could be made of substantive doctrines relating to emotions and the extent to which those doctrines, if genuinely inconsistent, ought to be reformed. We now take stock of what our inquiry has revealed.

Most of our analysis has been descriptive. We have attempted to show how two distinct conceptions of emotion that have long contended within the Western philosophical tradition are also at work, and often at war, within criminal law. These two views frequently inform competing doctrinal formulations, such as the common law and Model Penal Code versions of manslaughter, and the cognitive- and volitional-impairment tests for insanity. The interplay between them can also be seen in the curious history of many doctrines such as "premeditation" and "irresistible impulse," which clearly began as mechanistic concepts but which ultimately evolved into more evaluative ones. Finally, the evaluative view, in particular, helps us to understand the shifting content of the law's assessment of which emotional motivations are "reasonable": as the social norms that construct good character change, so too do the assessments that law makes of emotions that comport with or defy those norms.

---

440. See, e.g., *United States v. Clark*, 8 F.3d 839, 845–46 (D.C. Cir. 1993) (psychological trauma associated with childhood abuse can warrant downward departure from penalty indicated by Sentencing Guidelines in extraordinary cases); *United States v. Roe*, 976 F.2d 1216, 1218 (9th Cir. 1992) (same).

441. Cf. Cass R. Sunstein, *Legal Reasoning and Political Conflict* (forthcoming 1996) (defending "incomplete theorization" as politically and morally acceptable resolution of contentious issues).

Indeed, once the influence of the evaluative conception is revealed, it is possible to view the law as less confused than it might initially appear. Often what seems to be inconsistency on the part of the law is actually inattention on the part of observers to the law's evaluative mode of assessment. It isn't contradictory or confused, for example, for the law to punish the homophobe who kills a gay man more severely than the cuckold who kills his wife's paramour. If the law is prepared to treat the low valuation of the victim's life expressed in the homophobe's hatred as false, and the high valuation of fidelity or honor expressed in the cuckold's anger as true, it needn't either mitigate the punishment for both on voluntarist grounds or enhance it for both on consequentialist ones.

But exposing the two conceptions of emotion in criminal law does not acquit it of all confusion. It is quite hard, for example, to make complete sense of the prevailing, fictional concept of "premeditation": the evaluative view explains why courts don't take the concept literally, but it doesn't explain why the fiction itself exists or has persisted. Moreover, the law often appears to shift opportunistically between the mechanistic and evaluative views—condemning in one moment the battered woman who was not impelled to kill by fear, and yet excusing in the next the "true man" who chose to stand his ground and fight rather than endure the shame of flight. The best that can be said is that the law is genuinely ambivalent about the significance of emotions; it hasn't fully made up its mind between the mechanistic and evaluative views.

This ambivalence is counterproductive. Our primary normative claim in this Article has been that the law would be better if, in all instances, it were expressly evaluative. Doctrines that openly appraise offenders' emotions better serve all the recognized purposes of criminal law. They facilitate the accurate expression of society's moral condemnation, which is often calibrated to the quality of the valuation embodied in offenders' emotional motivations. They reinforce deterrence by supplying an acceptable theory of value for identifying preferred states of affairs and by inculcating desirable emotional dispositions. And they implement our considered judgments about individual desert, which turn just as much on the quality of a person's character as on the quality of her choices.

Most important of all, evaluative doctrines are superior to mechanistic ones because they are *honest*. We have shown that terms like "premeditation," "impulse," and "voluntariness" frequently obscure the substance of evaluative appraisals. Such misdirection is worthy of reform, moreover, not because transparency in law is, as an abstract matter, always preferable to obscurity,<sup>442</sup> but because in this particular setting obscurity produces bad results. It's when the law falsely denies its evaluative

---

442. See generally Meir Dan-Cohen, *Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 Harv. L. Rev. 625 (1984) (examining and defending selective obscurity in criminal law).

underpinnings that it is most likely to be incoherent and inconsistent; it is when the law refuses to take responsibility for its most contentious choices that its decisionmakers are spared the need to be principled, and the public the opportunity to see correctable injustice.

But if the law is made expressly and uncompromisingly evaluative, will it always be just? Obviously, the answer is no. It would be naive to pretend that decisionmakers will do what is right merely because we command them to be self-conscious. For one thing, the principles to which they self-consciously appeal may themselves be unjust. It isn't naive, however, to believe that a society will come closer to being the best it can be if it denies its courts and juries doctrinal cover for indulging their prejudices.

The evaluative view, in short, cannot assure justice. No *theory* can. But the evaluative view can push society toward justice by forcing it always to hear, if not to heed, the voice of its own conscience.