

10-2003

## Fundamental Retribution Error: Criminal Justice and the Social Psychology of Blame

Donald A. Dripps

Follow this and additional works at: <https://scholarship.law.vanderbilt.edu/vlr>



Part of the [Criminal Law Commons](#), and the [Law and Psychology Commons](#)

---

### Recommended Citation

Donald A. Dripps, *Fundamental Retribution Error: Criminal Justice and the Social Psychology of Blame*, 56 *Vanderbilt Law Review* 1383 (2019)

Available at: <https://scholarship.law.vanderbilt.edu/vlr/vol56/iss5/2>

This Article is brought to you for free and open access by Scholarship@Vanderbilt Law. It has been accepted for inclusion in *Vanderbilt Law Review* by an authorized editor of Scholarship@Vanderbilt Law. For more information, please contact [mark.j.williams@vanderbilt.edu](mailto:mark.j.williams@vanderbilt.edu).

# Fundamental Retribution Error: Criminal Justice and the Social Psychology of Blame

*Donald A. Dripps\**

I.	INTRODUCTION: OVERVIEW, QUALIFICATIONS, AND ROADMAP .....	1385
	A. <i>Overview</i> .....	1385
	B. <i>Qualifications</i> .....	1391
	C. <i>Roadmap</i> .....	1392
II.	PSYCHOLOGICAL RESEARCH FINDINGS ON THE SIGNIFICANCE OF PERSONAL AND SITUATIONAL FACTORS IN EXPLAINING EVENTS .....	1393
	A. <i>Personality and Situation as Sources of Human Behavior</i> .....	1393
	B. <i>Fundamental Attribution Error</i> .....	1396
III.	EXPLAINING THE LAW'S EMPHASIS ON ACTUAL RESULTS...	1401
IV.	EXPLAINING THE DEFENSES .....	1406
	A. <i>Personal and Impersonal External Causes: Insanity, Necessity, and Intoxication Contrasted With Duress, and the Abuse Excuse</i> .....	1406
	B. <i>Causing the Condition of One's Own Defense: Forfeiture Contrasted With Culpability Analysis</i> ..	1413
	C. <i>Mitigating Circumstances in Capital Cases</i> .....	1414
V.	NORMATIVE IMPLICATIONS .....	1416
	A. <i>Reducing Irrationality</i> .....	1416
	1. <i>Reconsidering the Significance of Resulting Harm</i> .....	1416
	2. <i>Distinguishing Situational, Third-Party, and Blame-the-Victim Defenses</i> .....	1416
	3. <i>Rules Versus Standards</i> .....	1417
	4. <i>The Trial Level</i> .....	1418

---

\* James Annenberg Levee Professor of Law and Criminal Procedure, University of Minnesota Law School. I am grateful to my colleague Gene Borgida, and to the participants at a faculty workshop at the University of San Diego Law School, for extremely useful comments on a prior draft. Errors of thought and expression are solely the author's.

B.	<i>Social Cognition and Theories of Punishment: Is There a "Fundamental Retribution Error"?</i> .....	1421
1.	Fundamental Attribution Error and the Judgments Required by Retributive Theory of Punishment .....	1424
a.	<i>Types of Retributive Theories</i> .....	1424
b.	<i>Blameworthiness and FAE</i> .....	1426
2.	Virtue-Based Theories of Punishment .....	1430
3.	Utilitarian Accounts .....	1433
VI.	ANTICIPATING TWO PLAUSIBLE CRITICISMS .....	1434
VII.	CONCLUSION .....	1436

At least since the M'Naghten case of the 1840s,<sup>1</sup> Anglo-American criminal law has concerned itself closely, famously, and contentiously with the psychology of the accused.<sup>2</sup> Another significant body of scholarship addresses the psychology of juries,<sup>3</sup> and other valuable research has approached some of the rules of criminal evidence from the perspective of social and cognitive psychology.<sup>4</sup> There has, however, yet to be a general investigation of what social

---

1. Daniel M'Naghten's Case, 8 Eng. Rep. 718 (H.L. 1843) (holding that the insanity defense excuses crimes committed when defendant, by reason of mental disease or defect, did not know the nature and quality of the act or did not know that act was wrong).

2. See, e.g., ABRAHAM S. GOLDSTEIN, *THE INSANITY DEFENSE* (1967); PETER W. LOW ET AL., *THE TRIAL OF JOHN W. HINCKLEY, JR.: A CASE STUDY IN THE INSANITY DEFENSE* (1986); MICHAEL MOORE, *LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP* (1983); NORVAL MORRIS, *MADNESS AND THE CRIMINAL LAW* (1982); STEPHEN J. MORSE, *THE JURISPRUDENCE OF CRAZINESS* (1986).

3. See, e.g., NEAL FEIGENSON, *LEGAL BLAME: HOW JURORS THINK AND TALK ABOUT ACCIDENTS* (2000); NORMAN J. FINKEL, *COMMONSENSE JUSTICE: JURORS' NOTIONS OF THE LAW* (1995); see also Dennis J. Devine et al., *Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups*, 7 J. PSYCHOL. PUB. POL'Y & L. 622 (2001) (reviewing studies).

4. See Susan Marlene Davies, *Evidence of Character to Prove Conduct: A Reassessment of Relevancy*, 27 CRIM. L. BULL. 504, 526-33 (1991) (rejecting Fundamental Attribution Error, or "FAE", as justification for exclusion of character evidence to show subsequent conduct); Chris William Sanchirico, *Character Evidence and the Object of Trial*, 101 COLUM. L. REV. 1227, 1242-46 (2001) (same); see also Victor J. Gold, *Federal Rule of Evidence 403: Observations on the Nature of Unfairly Prejudicial Evidence*, 58 WASH. L. REV. 497, 521-23 (1983) (suggesting FAE as justification for exclusion of character evidence); Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CAL. L. REV. 1051, 1087 (2000) (same); Miguel A. Mendez, *The Law of Evidence and the Search for a Stable Personality*, 45 EMORY L.J. 221, 236 (1996) (suggesting that psychological research might support admission through expert testimony of more specifically-defined behavioral dispositions); Christopher Slobogin, *Race-Based Defenses—The Insights of Traditional Analysis*, 54 ARK. L. REV. 739, 772 n.152 (2002) (noting that expert testimony on race-based defenses might help to overcome FAE). For a general discussion of FAE, see *infra* notes 20, 34-47 and accompanying text.

cognition research might teach us about the criminal law's pervasive concern with blameworthiness.<sup>5</sup>

This Article undertakes that investigation. It brings research on the psychology of social cognition to bear on the decision-making processes of public officials charged with the administration of criminal justice. The psychological research suggests that these decision makers, like most other human beings, are likely to overestimate the causal significance of personal choice, and to correspondingly underestimate the causal significance of situational factors in the behavior of others.<sup>6</sup> My thesis is that this observer's tendency to attribute conduct and its consequences to personality, rather than to situation, has important and disturbing implications for the theory and practice of criminal law.

## I. INTRODUCTION: OVERVIEW, QUALIFICATIONS, AND ROADMAP

### A. Overview

The idea of culpability or blameworthiness plays a central role in both the theory and practice of criminal justice. We have it on high authority that

---

5. Andrew Lelling has suggested that the psychological research on situation and personality in determining behavior weighs strongly against character-based theories of culpability. See generally Andrew E. Lelling, *A Psychological Critique of Character-Based Theories of Criminal Excuse*, 49 SYRACUSE L. REV. 35 (1998). Gilbert Harman has made a similar argument against virtue ethics generally. See generally Gilbert Harman, *The Nonexistence of Character Traits*, 100 PROC. OF THE ARISTOTELIAN SOC'Y 223 (2000). I am not convinced that the behavioral research on situation and personality really excludes the existence of character traits. See *infra* note 33 and accompanying text. For a defense of virtue ethics along these lines, see Gopal Sreenivasan, *Errors About Errors: Virtue Theory and Trait Attribution*, 111 MIND 47 (2002).

My primary point is cognitive, not behavioral, and, although related, is different and more powerful. Whatever the role of personality in determining behavior, observers tend to exaggerate it. Traditional virtue theory, for example, depends not only on the existence of character traits, but also on the ability of observers to identify and evaluate character traits. I do not believe that virtue ethics, conceived of in terms of character traits, can be rescued from the implications of FAE. It would follow that a virtue theory of punishment, such as that defended in Kyron Huigens, *The Dead End of Deterrence, and Beyond*, 41 WM. & MARY L. REV. 943 (2000), is suspect.

Harman does make this cognitive point about virtue ethics, but his discussion is terse and highly general. See generally Gilbert Harman, *Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error*, 99 PROC. OF THE ARISTOTELIAN SOC'Y 315 (1999). He confines the implications of FAE to a critique of virtue ethics, when it suggests much wider skepticism about blame and punishment generally, and about criminal justice more particularly. *Id.*

6. See *infra* text accompanying notes 41-65.

[t]he contention that an injury can amount to a crime only when inflicted by intention is no provincial or transient notion. It is as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil.<sup>7</sup>

There could be no crime, said Blackstone, without a “vicious will.”<sup>8</sup> Despite some exceptions, such as strict liability offenses, the principle that blameworthiness is a distinctive feature of criminal conduct remains fundamental in our legal culture.

Yet many features of prevailing criminal law are hard to reconcile with this principle of culpability. For example, a failed attempt typically is punished more leniently than a successful one, even though those engaging in failed attempts are no less culpable (or dangerous) than those who are successful.<sup>9</sup> The felony murder doctrine departs from the culpability principle in the opposite direction by imposing liability for murder on account of unintended killings, even in the absence of culpable recklessness.<sup>10</sup> The *Pinkerton* doctrine, in similar fashion, imposes accomplice liability on conspirators for all offenses foreseeably committed in furtherance of the conspiracy, even though the individual defendant had no knowledge that such offenses were planned or committed.<sup>11</sup> In each of these instances, the liability of the actor is determined more by moral luck than by the degree of deliberate wrongdoing.<sup>12</sup>

Moreover, the scope of affirmative defenses does not, to put it mildly, track the culpability principle very consistently. One who kills in the insane delusion that only death can save the victim from eternal damnation,<sup>13</sup> or one who kills when intoxicated beyond self-

---

7. *Morissette v. United States*, 342 U.S. 246, 250 (1952).

8. 4 WILLIAM BLACKSTONE, COMMENTARIES \*21.

9. See, e.g., WAYNE R. LAFAVE, CRIMINAL LAW § 6.4, at 567 (3d ed. 2000) (“As to statutory provisions concerning the sentences which may be imposed for all or a broad class of attempts, the most common in the modern recodifications is that which declares the attempt to be a crime one degree below the object crime.”).

10. See *id.* at 671-72 (describing felony murder rule).

11. *Pinkerton v. United States*, 328 U.S. 640 (1946) (upholding convictions both for conspiracy and for substantive offenses committed after defendant was incarcerated); LAFAVE, *supra* note 9, § 6.8(a), at 633-35 (discussing *Pinkerton* rule).

12. “Moral luck” is a phrase used in moral philosophy to refer to the tension between the proposition that moral standing should not depend on luck and the widespread intuition that luck does indeed influence moral standing. See, e.g., Andrew Latus, *Moral Luck*, THE INTERNET ENCYCLOPEDIA OF PHILOSOPHY, at <http://www.utm.edu/research/iep/m/moralluc.htm> (last visited Oct. 26, 2003).

13. See Sherry F. Colb, *The Andrea Yates Verdict: A Nation in Denial About Mental Illness*, FindLaw’s Legal Commentary, at <http://writ.news.findlaw.com/colb/20020327.html> (March 27,

control (and in many jurisdictions even beyond awareness of his actions),<sup>14</sup> can be held just as liable as one who kills with sane and sober calculation. Yet for centuries the law has reduced murder to manslaughter when the killer is provoked by such affronts as a spouse's infidelity.<sup>15</sup> More recently, courts have shown some receptiveness to defenses based on wrongs done by the victim to the accused—the now-familiar claim of an “abuse excuse.”<sup>16</sup> The entrapment defense, in all its forms, exculpates blameworthy actors lured into committing a crime by government agents.<sup>17</sup>

These anomalies are sometimes thought to reflect imperative needs for social control at the expense of fairness.<sup>18</sup> Sometimes they are thought to reflect normative disagreements about the degree of culpability properly attaching to various forms of conduct.<sup>19</sup> Without

---

2002) (discussing murder conviction of Andrea Yates, who drowned her children while mentally ill in the belief that only death could save their souls). An insane belief that God has commanded the specific crime can exculpate the accused under the so-called “deific decree” doctrine, but the doctrine is understood as a narrow exception to the usual limits of the insanity defense. *Schmidt v. State*, 110 N.E. 945, 949 (N.Y. 1915) (Cardozo, J.) (recognizing deific decree doctrine in dictum); Grant H. Morris & Ansar Haroun, “*God Told Me to Kill: Religion or Delusion?*,” 38 SAN DIEGO L. REV. 973 (2001) (surveying deific decree cases).

14. A substantial minority of American jurisdictions hold the defendant responsible even when, on account of voluntary intoxication, he lacks the intent or knowledge required by the definition of the charged offense. *Montana v. Egelhoff*, 518 U.S. 37, 48 n.2 (plurality opinion) (1996) (noting that ten states do not permit intoxication to negative intent or knowledge). The majority of jurisdictions, moreover, deny any defense based on intoxication when the charged crime requires recklessness or negligence. See LAFAVE, *supra* note 9, § 4.10(c), at 416.

15. See, e.g., LAFAVE, *supra* note 9, § 7.10(b), at 707 (“It is the law practically everywhere that a husband who discovers his wife in the act of committing adultery is reasonably provoked, so that when, in his passion, he intentionally kills either his wife or her lover (or both), his crime is voluntary manslaughter rather than murder.” (footnote omitted)).

16. See, e.g., ALAN M. DERSHOWITZ, *THE ABUSE EXCUSE* (1994) (attaching “abuse excuse” label to use of expert testimony calculated to shift responsibility from the accused to her or his past abuses and strongly criticizing this practice).

17. See, e.g., *Jacobson v. United States*, 503 U.S. 540, 550-51 (1992) (holding that, under subjective federal entrapment defense, proof that government induced offense is not rebutted by proof that defendant engaged in similar conduct at time when conduct was not illegal). What result if Jacobson had claimed that his possession of child pornography was attributable to sexual obsession rather than government temptation? Lacking some human agent to shoulder the blame, he would not even have been allowed to present this theory to the jury.

18. As Professor Morse says, discussing *Egelhoff*, “[c]oncern with culpability thus almost always conflicts with concern for public safety.” Stephen J. Morse, *Fear of Danger, Flight from Culpability*, 4 PSYCHOL. PUB. POL’Y & L. 250, 256 (1998).

19. There is, for example, a lively debate about whether provocations such as spousal infidelity or homosexual advance reduce the culpability of killings, as supposed by the voluntary manslaughter category. See, e.g., Joshua Dressler, *Why Keep the Provocation Defense?: Some Reflections on a Difficult Subject*, 86 MINN. L. REV. 959 (2002); Joshua Dressler, *When “Heterosexual” Men Kill “Homosexual” Men: Reflections on Provocation Law, Sexual Advances, and the “Reasonable Man” Standard*, 85 J. CRIM. L. & CRIMINOLOGY 726 (1995); Emily L. Miller, *(Wo)manslaughter: Voluntary Manslaughter, Gender, and the Model Penal Code*, 50 EMORY L.J.

purporting to reject such accounts completely, this Article explores a different kind of explanation—one rooted in modern research findings of social psychology.

Psychological research indicates that observers tend to hold the individual, rather than the situation, responsible for the individual's action.<sup>20</sup> Although people tend to attribute their own misconduct to external constraints, they tend to attribute the behavior of others to personality rather than context. Psychologists regard this tendency—variously termed the correspondence bias or fundamental attribution error (henceforth FAE)—as firmly established.<sup>21</sup> It may not prevail, or prevail only to a lesser extent, in non-Western cultures. There is no consensus about what causes FAE. But about the basic phenomenon there is little doubt. It is commonsensical for observers to conclude that other individuals chose as they did, not because of the situations those individuals were in, but because of the kind of people they are. Common sense, however, turns out to be wrong.

As a descriptive matter, FAE helps to explain many puzzling features of existing criminal law. For example, theorists have experienced considerable difficulty explaining the more lenient treatment accorded failed as compared to successful attempts. The distinction may reflect a cognitive tendency rather than a normative anomaly; decision makers may irrationally but consistently attribute failure and success to the actor rather than to the circumstances. The same type of tendency may in part explain both the felony murder doctrine and the expansive view taken in some of the proximate cause cases.

---

665 (2001); Robert B. Mison, *Homophobia in Manslaughter: The Homosexual Advance as Insufficient Provocation*, 80 CAL. L. REV. 133 (1992).

20. See *infra* notes 41-65 and accompanying text.

21. See *infra* note 41 and accompanying text. Terminology such as “bias” and “error” implies deviation from an identifiable norm of rationality. It is by no means clear that social cognition really departs from such an identifiable norm. See, e.g., Gregory Mitchell, *Taking Behavioralism Too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law*, 43 WM. & MARY L. REV. 1907, 1943 (2002). It may be more precise to speak of cognitive tendencies rather than biases, or of information-processing strategies rather than errors. The foundational literature in this area, however, uses the “fundamental attribution error” label, and that nomenclature has become so settled as to render other terminology potentially confusing. In the paper that follows, I shall use FAE as a term of art to refer to the demonstrated tendency of observers to attribute conduct and consequences to person rather than situation. In many places, jurists have failed to articulate a rational justification for legal doctrines that can be explained as products of overassessing personal as distinct from situational factors in the stream of events that led to a negative outcome. When generations of scholars have tried and failed to identify a convincing rational justification for such doctrines, it seems fair to regard the decision process that produced them not only as partly determined by cognitive predispositions, but as erroneously determined to that extent.

F AE also suggests, contrary to the popular impression that abuse excuses have undermined the law's assumption of free will, that legislators, courts, and juries will likely define the grounds of excuse too narrowly when the claimed excuse is rooted in purely situational factors rather than third-party wrongdoing. The legal indifference to the defendant's intoxication,<sup>22</sup> the widespread doctrine that reckless or negligent mistake about the facts supporting a defense bars that defense completely,<sup>23</sup> and the abandonment of the behavioral prong of the insanity defense,<sup>24</sup> all reflect the judgment that the individual, rather than the situation, caused the harm the law aims to prevent. Social and cognitive psychology teaches us to distrust that judgment. Each of these doctrines should be reconsidered with a healthy dose of skepticism about intuitive judgments regarding personal responsibility.

When, by contrast, the defendant invites the court to blame another individual, the psychological dynamic changes. The provocation defense to murder is an ancient example; the abuse excuse is a more modern counterpart. Entrapment, too, involves wrongdoing by players other than the accused. In rape cases, the jury may be encouraged to blame the victim for sexual activity. This may very well translate into an irrational inference of consent from the victim's character.

As a normative matter, FAE has two distinct and largely alternative implications. The first is that FAE should be recognized as a standing risk in legal decision making. Policy analysts should take a hard look at both prevailing doctrine and proposed reforms with some appreciation of the cognitive tendency to hold persons rather than situations accountable for bad outcomes. Rule-drafters and judges should handle trial procedures with a similar awareness.

The second implication is that if, as seems not unlikely, cognitive defects are relatively immune to rational correction, the old debate over retributive and utilitarian theories of punishment should be seen as both more complicated and more important than hitherto.<sup>25</sup> Efforts to institutionalize retributive theory are complicated by the standing risk, reflected in some strands of prevailing law, of overassessing blameworthiness. Unchecked, FAE may mean that retributive theory in practice will inflict punishment out of proportion to a rational measurement of just deserts. Legislators, judges, and

---

22. See *infra* text accompanying note 121-122.

23. See *infra* text accompanying notes 123-125.

24. See *infra* text accompanying notes 99-112.

25. For general treatments of the philosophy of punishment, see *infra* note 151.

juries following intuitive notions of blameworthiness will tend to overassess individual responsibility and underassess situational factors. It follows that a system based on retributivism will, in practice, tend to overpunish, even according to retributive criteria. To the extent that utilitarianism relies on common sense assessments of blameworthiness to limit punishment, utilitarians too must grapple with the challenge posed by FAE.

FAE's implications for the theoretical debate, however, are not entirely one-sided. Positive or mandatory retributivist theorists, who believe that culpable wrongdoing is a necessary and sufficient reason for punishment, may well have second thoughts about institutionalizing a theory that requires punishment according to what in practice will be exaggerated assessments of blameworthiness. Permissive retributivists should be less discomfited, because they believe that blameworthiness only authorizes but does not require punishment.<sup>26</sup> According to this view, exaggerated assessments of blameworthiness would translate into actual punishment only if pragmatic considerations counseled punishing to the limits set by those assessments. This would still be a disturbing risk, given the priority retributivists typically place on avoiding undeserved punishment.

The utilitarian theorist is in much the same boat as the permissive retributivist. Typical utilitarian theories accept the idea that culpability is a necessary condition of punishment. The utilitarians recognize culpability-based side-constraints because otherwise their theory might justify punishment of the innocent, recognition of strict liability, imposition of cruel punishments, and indeed anything else that would deliver a favorable balance of consequences. If decision makers tend to exaggerate culpability, however, the utilitarian, like the permissive retributivist, runs a standing risk of punishing in excess of what her side-constraints would permit if rationally applied.

Not only does the debate become more complicated, it also becomes more important. If it is possible to reduce the influence of FAE, retributivists could conceivably find themselves in dogged disagreement with their utilitarian colleagues. If the level of punishment now prevalent is justified by a favorable balance of consequences, and current assessments of culpability are overstated, retributive theory would require reductions in punishment that are inconsistent with social welfare. The comforting consistency of

---

26. For more on the distinction between permissive and mandatory retributivism, see *infra* note 156.

utilitarian and retributive theories with respect to most practical issues may be a cognitive illusion.

### B. Qualifications

Before proceeding to substantiate the argument, I would like to distinguish four propositions I regard as mistaken, but which might easily be confused with the position I am defending. I am not, in the first place, making a relativistic or deterministic claim that no one can be justly blamed for anything.<sup>27</sup> My claim is not that personality never determines behavior, but only that observers are likely to overestimate the extent to which this is the case.

Second, I am not making an elitist claim that decisions about responsibility should be made by psychologists rather than jurors, or by judges rather than legislators.<sup>28</sup> The correspondence bias does not disappear in professional bureaucrats.<sup>29</sup> What the law needs to do is

---

27. For the manifesto of pure behaviorism, see B.F. Skinner, *BEYOND FREEDOM AND DIGNITY* (1971).

28. One persistent theme in Professor Finkel's work is that "law" and "commonsense justice" diverge and that "commonsense justice" often has the advantage over "law." See, e.g., FINKEL, *supra* note 3, at 319-37. Jurors, however, are surely subject to cognitive predispositions such as FAE, as Professor Finkel has recognized. See, e.g., Norman J. Finkel, *Commonsense Justice, Psychology, and the Law: Prototypes that are Common, Senseful, and Not*, 3 *PSYCHOL. PUB. POLY & L.* 461, 477-78 (1997) (noting that capital-sentencing procedures may exacerbate rather than reduce FAE). Finkel's point is less that jurors are better than other decision makers than that they are at least no worse. "If people have a dispositional nature, bias, or penchant and the social psychological evidence for this is persuasive, in my opinion, then, social scientists are not exempt." *Id.*

Two features of Finkel's work call for comment here. First, he understands "the law" as exogenous, rather like an inscribed totem that came from nowhere. The enacted law, however, is made, by legislators with an eye on the same community that generates Finkel's "commonsense justice," and by judges both appointed and elected. The doctrinal features Finkel finds arbitrary or random—felony murder, provocation, and insanity—may very well be the systematic products of cognitive bias operating on legislators, judges, and voters. See FINKEL, *supra* note 3, at 154-71 (discussing felony murder); *id.* at 297-318 (discussing provocation); *id.* at 261-297 (discussing insanity).

Second, Finkel pays little or no attention to those cases in which community sentiment agrees with the enacted law. The fact is that a conviction can be upheld on appeal only if the jury and the judges agree. The reported cases therefore provide considerable evidence that juries are not wildly out of step with the positive law.

29. For example, in their study of popular attitudes toward prevailing legal doctrine, Robinson and Darley found that the subjects opposed equating the penalties for attempted and completed offenses, but also were sympathetic to treating killings during felonies as manslaughter rather than murder. PAUL H. ROBINSON & JOHN M. DARLEY, *JUSTICE, LIABILITY, AND BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW* 27 (1995) ("In the view of our subjects . . . the level of punishment for attempt ought to be significantly less than that for the completed offense."); *id.* at 180 ("But according to our subjects, current doctrine goes too far, for it punishes such a negligent killing as if it were murder, although the subjects would prefer to punish it as manslaughter."). In the case of punishment for attempts, popular opinion can be

not change the decision makers, but account for the risk that decision makers will overestimate personal as opposed to situational factors.

Third, FAE is not a phenomenon that inevitably disfavors the accused. For example, the exaggeration of personal, relative to situational, factors helps to explain the lenience accorded failed attempts. Affirmative defenses that cast blame on the victim or the police offer another pro-defendant example. Defenses such as provocation and entrapment may derive some of their popularity from the tendency to exaggerate personal, as distinct from situational, explanations for behavior and its consequences. My thesis holds that the law should recognize FAE as a systemic tendency to find *someone* responsible when harm occurs. Although that someone will generally be the accused, this is by no means invariably the case.

Finally, attributing conduct and results to personality rather than situation is not inevitable. Overassessing personal responsibility is a risk, not a certainty. Some of the doctrinal rules that FAE might help to explain are in fact minority positions in American jurisdictions. The doctrinal anomalies FAE helps to explain, moreover, may also derive support from other considerations, some perfectly rational. Thus, recognizing our common tendency in favor of personal responsibility would not automatically translate into doctrinal change. We would, however, analyze doctrine with greater skepticism about assessments of personal responsibility.

### C. Roadmap

Part II summarizes the relevant research findings in social psychology. Parts III and IV suggest that some puzzling features of current doctrine may be due in part to pervasive cognitive tendencies. Part III examines the "moral luck" problem in the contexts of attempt, felony murder, and conspiracy, and argues that FAE may explain the law's preoccupation with harm actually caused rather than with dangerousness or wickedness manifested by conduct. Part IV analyzes the defenses, and suggests that FAE may help to explain the law's more generous response to defenses based on victim misconduct than to those based on situational factors such as insanity and intoxication. It may well be that most novel claims of situational excuse deserve to be rejected; ultimately, each must be judged on its own merits. In the process of judging, however, we should do what we can to account for

---

interpreted as attributing outcome (the failed attempt) to personality rather than situation, as predicted by research on FAE. In the case of felony murder, however, it is popular opinion that inclines toward mitigating the current felony murder rule, which attributes outcomes to personality automatically without any inquiry into *mens rea*.

our tendency to focus on the defendant's behavior rather than on the situation.

Part V takes up the normative issues related to FAE. The criminal justice system should do what it can to recognize and resist FAE. It seems likely, however, that the cognitive tendencies at work cannot be completely dispelled. In that case, the psychological research on attribution has normative implications for criminal law theory's concern with blameworthiness. Agency, responsibility, and causation are the wellsprings of blameworthiness; the psychological research indicates that people are prone to overestimate agency, responsibility, and causation. Part V explores these implications, expressing some skepticism about theories that require, rather than permit, punishment according to desert. Part VI briefly responds to two plausible criticisms: defense of the rational-actor assumption and reservations predicated on the risks attending law-office psychology.

The Article aims primarily at explaining otherwise puzzling divergences between prevailing theory and prevailing practice. For the most part, specific doctrinal recommendations are not advanced. My ultimate claim, then, is neither for nor against specific rules or procedures, but rather a plea for broader awareness that those assessing criminal responsibility are, like those they judge, human—all too human.

## II. PSYCHOLOGICAL RESEARCH FINDINGS ON THE SIGNIFICANCE OF PERSONAL AND SITUATIONAL FACTORS IN EXPLAINING EVENTS

### *A. Personality and Situation as Sources of Human Behavior*

Human behavior, like other phenomena, is subject to causation. A rich psychological literature addresses the respective causal power of individual personality and impersonal situations on the resulting behavior. As we shall see, the current state of the research is, for criminal law purposes, less important than it might seem. But, the personality-situation literature very helpfully frames the issue posed to criminal law by FAE.

People very commonly attribute conduct to character traits. We blame the fight on the bully, the accident on the klutz, the divorce on the skirt-chaser, and so on. Ordinary language includes thousands of words for character traits, which suggests just how commonly

people reason from individual proclivities to actual behavior.<sup>30</sup> Early in the twentieth century, leading psychologists accepted trait theory as a valid description of human behavior.<sup>31</sup> Their efforts to confirm the predictive power of character traits, however, had the opposite effect.<sup>32</sup> Situation may not be the exclusive determinant of behavior, but the causal significance of character traits defined as generally as “pugnacious” or “dishonest” is much weaker than commonsense trait theory suggests.<sup>33</sup>

The psychological research on behavior supports three basic findings. First, individual responses to specific situations are reasonably consistent across time.<sup>34</sup> If subject A cheats on an examination given one opportunity, A is likely to cheat on a similar exam if given another opportunity.<sup>35</sup> This finding, obviously, doesn't contradict trait theory.

The second finding is that individual behavior is *not* consistent across *situations*.<sup>36</sup> The fact that A cheats on an examination does not

30. See G.W. Allport & H.S. Odbert, *Traitnames: A Psycho-lexical Study*, 47 PSYCHOL. MONOGRAPHS 171-220 (1936).

31. See LEE ROSS & RICHARD E. NISBETT, *THE PERSON AND THE SITUATION* 90-118 (1991).

32. *Id.*

33. *Id.*

34. See, e.g., ZIVA KUNDA, *SOCIAL COGNITION* 417-18 (1999).

Yet the handful of studies that have risen to the challenge over the last seven decades or so have yielded remarkably similar conclusions: People often behave quite consistently when they encounter the same situation at different times, but there is very little consistency in their trait-related behavior from one situation to other, different situations. More technically, the temporal stability of trait-related behaviors is high, but their cross-situational consistency is low.

*Id.* (citation omitted).

35. See, e.g., ROSS & NISBETT, *supra* note 31, at 101 (1991).

Stability coefficients—the correlation between two measures of the same behavior on different occasions—often exceed .40, sometimes reaching much higher. For example Hartshorne and May (1928) found that the tendency to copy from an answer key on a general information test on one occasion was correlated .79 with copying from an answer key on a similar test six months later. Newcomb (1929) found that talkativeness at lunch was a highly stable attribute; it just wasn't very highly correlated with talkativeness on other occasions (see also Buss & Craik 1983, 1984). [Walter] Mischel insisted that strong differences between people were apparently limited to specific responses to specific situations, for example, friendliness in the lunchroom or willingness to confront one's employer, not broad, cross-situational, extroversion or assertiveness.

*Id.*

36. See generally WALTER W. MISCHEL, *PERSONALITY ASSESSMENT* (1968); DONALD R. PETERSON, *THE CLINICAL STUDY OF SOCIAL BEHAVIOR* (1968). Ross and Nisbett summarize the research results as follows:

justify a prediction that A will steal unguarded property. It has been suggested that, given a large sample of prior behavior and appropriately qualified definitions of the relevant traits, personality can be used to predict behavior.<sup>37</sup> It still seems fair to say that the traits described by ordinary language—dishonesty, pugnacity, loquaciousness—describe behavioral propensities too broadly. Our intuitions about general character and behavior are, shortly put, incorrect.

The third finding is that individual responses to specific situations are surprisingly consistent across persons. In the notorious Milgram experiments, for instance, sixty-eight percent of the subjects inflicted what they were led to believe were extremely painful and dangerous electric shocks.<sup>38</sup> The Stanford prison experiment, conducted in the 1970's, similarly showed that an extreme environment could produce antisocial behavior in ordinary subjects.<sup>39</sup> It is comforting to think that one's self could never become a murderer, a torturer, a rapist, or a drug dealer. No doubt there are persons whose internal controls dominate situational factors. The empirical evidence, however, seems to suggest that such highly ethical individuals are the exception rather than the norm.

Nonetheless, both personality and situation influence behavior. If (for instance) a trait and a behavioral response to a situation have a correlation coefficient of .20, seventy percent of those with the trait will engage in the behavior while only fifty percent of those without the trait will engage in the behavior. The ascendant view in psychology is now "interactionism," and holds unsurprisingly that both personality and situation play important roles in causing human behavior.<sup>40</sup>

---

the average correlation between different behavioral measures specifically designed to tap the same personality trait (for example, impulsivity, honesty, dependency, and the like) was typically in the range between .10 and .20, and often was even lower. . . . Correlations between scores on personality scales designed to measure a given trait and behavior in any particular situations presumed to tap that trait rarely exceeded the .20 to .30 range.)

ROSS & NISBETT, *supra* note 31, at 95.

37. ROSS & NISBETT, *supra* note 31, at 107-17.

38. Stanley Milgram, *Behavioral Study of Obedience*, 67 J. ABNORMAL & SOC. PSYCHOL. 371 (1963).

39. See Craig Haney et al., *Interpersonal Dynamics in a Simulated Prison*, 1 INT'L J. CRIMINOLOGY & PENOLOGY 69 (1973).

40. See, e.g., Lelling, *supra* note 5, at 86-87. Lelling notes the influence of both factors: Trait psychologists and situationists continue to debate the point, but by now there is general agreement within the field that while behavior is undeniably influenced by the surrounding environmental context, traits, that is, stable,

We do not need to make any generalized assessment of the relative causal power of personality and situation to understand the implications of FAE for criminal justice. Most of the time, the situational determinants of crime are not exculpatory. The purse-snatcher commits robbery only in response to an apparently vulnerable victim and only in the absence of the police. Such situational factors earn the purse-snatcher no moral points.

Not infrequently, however, the criminal defendant claims that he responded to a situation that does earn moral points for the defendant. He may claim, for example, that he behaved as he did in response to duress, insanity, intoxication, or entrapment. Typical people, such a defendant argues, do not commit robbery simply because they encounter a vulnerable victim in the absence of the police. But, typical people would commit crimes similar to those with which I am charged if they found themselves in the situation that confronted me. If the situation that caused the defendant's behavior is morally exculpatory, criminal justice decision makers must attribute the defendant's conduct and/or its consequences to the defendant himself or to the situation. Attribution to the defendant implies guilt; attribution to the exculpatory situational factors implies innocence or at least mitigation.

Human beings, however, are not very good at making such attributions. The next section takes up the research supporting this tendency to exaggerate personality relative to situation.

### *B. Fundamental Attribution Error*

The research just discussed recognizes that both personality and situation matter in determining behavior. Different people do respond differently to similar situations, although the differences in response appear to be less than that supposed by trait theory. Research on human cognition, however, adds another set of findings relevant to criminal law. The gist of these findings is that observers tend to attribute behavior and its consequences to personality rather than to situation.<sup>41</sup>

---

internal dispositions affecting behavior across different situations, do exist.

They just play a far smaller role in human behavior than generally believed.

*Id.* (footnotes omitted); accord Davies, *supra* note 4, at 518 ("The trait-versus-situation controversy in the field of personality and social psychology has produced widespread agreement that behavior is simultaneously a function of disposition and situation, and their mutual interaction." (footnote omitted)).

41. See, e.g., FEIGENSON, *supra* note 3, at 57-62 (2000) (discussing studies and considering possible sources of FAE); KUNDA, *supra* note 34, at 428-32 (discussing studies and concluding: "In sum, we make the fundamental attribution error, overattributing behavior to dispositions

In one now-classic experiment, the experimenters asked the subjects to identify the true opinion held by writers or speakers on controversial topics—in the original experiment, essays attacking or defending Fidel Castro.<sup>42</sup> When told that the essay writers had freely chosen their position, the subjects understandably attributed a pro-Castro view to the authors of the pro-Castro essays.<sup>43</sup> But when the subjects were told that the essay writers had been given no choice over the point of view they were assigned to defend, the subjects still tended to attribute the viewpoint expressed in the essay to the author.<sup>44</sup> The observers, in other words, ascribed behavior to the person even when they had been informed of a powerful situational constraint.

In another experiment, observers were asked to rate the intelligence of participants in a quiz game.<sup>45</sup> The observers tended to rate the person asking the questions as more intelligent than the players attempting to answer, even though participants were assigned roles randomly and even though it should have been obvious that the person asking the questions could exploit personal areas of specialized knowledge.<sup>46</sup> In still another experiment, subjects were asked to evaluate the abilities of two groups of basketball players assigned to shoot free throws, one group in a well-lit gymnasium and the other group in a dimly lit facility.<sup>47</sup> Naturally enough, the players in the good lighting made a higher percentage of shots than those in the more dimly lit gym; nonetheless, the observers tended to rank those players with higher shooting percentages as better players notwithstanding the situational difference.<sup>48</sup>

---

and underattributing it to situations because we often do not appreciate the power of situations.”); ROSS & NISBETT, *supra* note 31, at 125-33 (describing studies; concluding, at 133, “People readily make trait ascriptions from data that permit only a situational interpretation or, at most, the interpretation that the actor behaves in a particular way in a particular type of situation.”); Daniel T. Gilbert & Patrick S. Malone, *The Correspondence Bias*, 117 PSYCHOL. BULL. 21 (1995) (reviewing studies).

42. See Edward E. Jones & Victor A. Harris, *The Attribution of Attitudes*, 3 J. EXPERIMENTAL SOC. PSYCHOL. 1 (1967).

43. The subjects’ responses to the essays were rated on a scale from 10 (extremely anti-Castro) to 70 (extremely pro-Castro). *Id.* at 5. When told the authors chose their positions voluntarily, the mean attribution was 59.62 for pro-Castro essay writers and 17.38 for anti-Castro essay writers. *Id.* at 6. In the no-choice condition the attributions were 44.10 and 22.87 respectively. *Id.*

44. *Id.* For discussion, see ROSS & NISBETT, *supra* note 31, at 126.

45. See KUNDA, *supra* note 34, at 429-30.

46. Lee Ross et al., *Social Roles, Social Control, and Biases in Social-Perception Processes*, 35 J. PERSONALITY SOC. PSYCHOL. 485 (1977). For discussion, see KUNDA, *supra* note 34, at 428-30.

47. FEIGENSON, *supra* note 3, at 57.

48. *Id.*

Moreover, the size of the effect of FAE appears to be substantial. That is to say, not only do the studies show statistically significant results in the sense that chance is unlikely to explain the results, the studies also show that the magnitude by which observers exaggerate the causal power of personality is large.<sup>49</sup> In the original Fidel Castro experiment, the subjects' perceptions of the essay writers' attitudes toward Castro were measured on a scale from 10 (most anti-Castro) to 70 (most pro-Castro).<sup>50</sup> When observers were told that the writers chose their own position, the mean score those observers gave to the pro-Castro writers was 59.62 and to the anti-Castro writers was 17.38.<sup>51</sup> Those observers who were told that the positions had been assigned still produced mean scores of 44.10 for the pro-Castro writers and 22.87 for the anti-Castro writers.<sup>52</sup> "Perhaps the most striking result of the . . . experiment was the tendency to attribute correspondence between behavior and private attitude even when the direction of the essay was assigned."<sup>53</sup>

In the quiz master experiment, observers were asked to rate the general knowledge of the participants.<sup>54</sup> Since assignment to the roles of quiz master and contestant was random, one would have expected the ratings to be approximately equal. Observers, however, rated the questioners with a score almost 70% higher than the level of the contestants.<sup>55</sup> In another study, researchers asked the subjects to predict which of two persons would behave more honestly on a future occasion, given that one had behaved less honestly than the other on a prior occasion.<sup>56</sup> "The actual correctional between behaviors in any two situations tapping traits such as honesty or friendliness is at best .15, yet participants believed it to be close to .80."<sup>57</sup>

Actors have different cognitive predispositions than observers. Actors tend to attribute their failures to situational constraints rather than to personality, and their successes to personality rather than to the situation.<sup>58</sup> This tendency may not result solely from self-

---

49. For a discussion of the difference between statistical significance and effect size, see ROSS & NISBETT, *supra* note 31, at 21.

50. See Jones & Harris, *supra* note 42, at 5.

51. *Id.* at 6.

52. *Id.*

53. *Id.* at 7.

54. See KUNDA, *supra* note 34, at 430.

55. *Id.* (observers rated questioners at 82.9 and contestants at 48.9).

56. *Id.* at 433-34.

57. *Id.*

58. See ROSS & NISBETT, *supra* note 31, at 140-41 (reviewing studies, concluding "[r]ampant dispositionism is kept in check when it is the self that is in question."); Tom Pyszczynski & Jeff Greenberg, *Toward an Integration of Cognitive and Motivational Perspectives on Social*

sympathy. There is some evidence that the actor-observer differences result from the operation of an availability heuristic. The actor must both observe the situation and respond to it; the observer needs only observe.<sup>59</sup>

Even when people know that the behavior of others results from context rather than personal choice, they tend to attribute behavior to personality. This may, for instance, help to explain why readers may be skeptical of the situationist critique of trait theory. Even if trait theory turned out to have more power than current research supports, however, FAE would suggest that observers tend to overstate the degree to which personality traits determine behavior.

As Professor Feigenson has explained, FAE implies that observers tend not just to attribute behavior to personality, but also to attribute undesirable consequences to behavior rather than to circumstances.<sup>60</sup> Observers resist the conclusion that serious harm might have resulted from chance rather than from wrongdoing. When harm has occurred, this tendency is magnified by the desire to believe

---

*Inference: A Biased Hypothesis-Testing Model*, 20 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 297, 298 (1987) (noting a "well-replicated finding that people tend to make dispositional attributions for their successes and situational attributions for their failures").

59. ROSS & NISBETT, *supra* note 31, at 141. Ross and Nisbett describe observers' perceptions of the world by noting that "[p]eople are active, dynamic, and interesting; and these are the stimulus properties that direct attention. The situation, in contrast, is normally relatively static and known only hazily." *Id.* at 139.

60. FEIGENSON, *supra* note 3, at 59. Feigenson argues that

The fundamental attribution error should dispose jurors to find that accidents are due to someone's negligence, that "someone's to blame." Assume that no one is responsible for the accident, in the sense that no one is at fault. The accident "just happened." This amounts to saying, of each party, that there is no good reason to expect that party to have acted differently to avoid the accident. Under the circumstances, anyone else would have behaved the same way. The fundamental attribution error, by contrast, attributes the actor's behavior to his or her personality, to something "in" the actor, rather than to the circumstances. This is tantamount to saying that some substantial percentage of others would have acted differently under the circumstances. (If most others would have acted the same way as the actor did, then the attribution, logically, would have to be to the circumstances, because nothing "in" the actor led him or her to act differently from the norm.)

*Id.* Feigenson is talking about tort, not crime, but FAE's tendency to exclude claims of situational excuse and impose liability for bad moral luck would seem to transfer precisely to the criminal context.

that the world is basically just<sup>61</sup> and that unlikely events that in fact occur were more likely *ex ante* than in fact they were.<sup>62</sup>

Psychologists disagree about the underlying cause of FAE.<sup>63</sup> Evidence that FAE is less pronounced in non-Western cultures also clouds the picture.<sup>64</sup> For purposes of American criminal law, however, these intriguing issues need not divert us. We are concerned with decision makers in a Western legal system. If they are likely to overassess personal responsibility, we need to account for that tendency, whatever its cause and whether or not a similar predisposition operates in other cultures. I do not deny the universality of principles of justice. I claim only that efforts to institutionalize principles of justice should take due account of the cognitive tendencies of those who will operate the institutions.

If we put the situationist critique of trait theory together with the research on FAE, the resulting description of human behavior and cognition is thoroughly disturbing for anyone concerned with criminal justice. The law describes criminal behavior as the free choice of deviant individuals, when in fact human behavioral responses to specific situations vary less across persons than is commonly supposed.<sup>65</sup> Worse yet, because of FAE, observers tend to attribute conduct to personality even when faced with facts that clearly indicate that the behavior was controlled by the situation. Standard accounts of criminal justice call on observers to evaluate the personal responsibility of the accused for harm caused. Observers, however, will not be able to perform this task in a completely rational way. Quite the contrary; observers—whether legislators, police and

---

61. On the tendency to believe in a just world, see MELVIN J. LERNER, *BELIEF IN A JUST WORLD: A FUNDAMENTAL DELUSION* (1980). Given a just world, human suffering needs to be blamed on human wrongdoing—a psychic need that can be satisfied by blaming the victim instead of the perpetrator, but which in any event tends to minimize impersonal causes of negative events.

62. FEIGENSON, *supra* note 3, at 62 (“The hindsight bias is the tendency to overestimate the probability of a known outcome and the ability of decision makers to have foreseen it. . . . The hindsight bias is one of the most consistently replicated effects in the cognitive psychology literature and has proved fairly resistant to attempts to reduce its impact.” (citations omitted)).

63. See, e.g., KUNDA, *supra* note 34, at 441-43 (speculating about possible causes, including the focus of Western culture on personality traits, leading observers to view and remember others through the light of earlier experiences with those individuals, and the fact that while a given observer may see another individual in many different situations, all of those situations have in common the observer’s presence); ROSS & NISBETT, *supra* note 31, at 139-40 (speculating about possible causes).

64. See FEIGENSON, *supra* note 3, at 58 n.15 (“Cross-cultural research, moreover, indicates that the bias toward dispositional attributions may derive from broader cultural meaning systems, for example, the Western emphasis on individualism, rather than from more or less universal cognitive processes.” (citations omitted)).

65. See *supra* notes 30-39 and accompanying text.

prosecutors, judges, or jurors—will be predisposed toward attributing conduct and its consequences to personality.

### III. EXPLAINING THE LAW'S EMPHASIS ON ACTUAL RESULTS

Whether we regard the culpability principle as the reason for punishment, as retributivists do, or as a side-constraint on the pursuit of collective advantage, as utilitarians do, we will have difficulty explaining the various situations in which the positive criminal law varies the punishment for wrongdoing according to its actual consequences.<sup>66</sup> At least three prominent and heavily criticized doctrinal anomalies turn on such assessments of harm actually caused.

First, one who attempts but fails to commit an offense is generally subject to a lesser penalty than one who attempts and succeeds.<sup>67</sup> Sometimes the difference in punishment is dramatic; for example, no American jurisdiction authorizes the death penalty for attempted murder.<sup>68</sup> Yet it is hard to explain why one whose attempt

---

66. For some general treatments of the significance of resulting harm for purposes of evaluating blameworthiness, see, for example, Sanford H. Kadish, *Foreword: The Criminal Law and the Luck of the Draw*, 84 J. CRIM. L. & CRIMINOLOGY 679 (1994); Stephen J. Schulhofer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497 (1974). See generally Symposium, *Harm vs. Culpability: Which Should Be the Organizing Principle of the Criminal Law?*, 5 J. CONTEMP. LEGAL ISSUES 1 (1994).

67. See, e.g., LAFAVE, *supra* note 9, § 6.4, at 567.

Considerable variation is to be found across the country concerning the authorized penalties for attempt. As to statutory provisions concerning the sentences which may be imposed for all or a broad range of attempts, the most common in the modern recodifications is that which declares the attempt to be a crime one degree below the object crime. Another common provision establishes categories according to the severity of the penalty for the completed crime and specifies a range of penalties for attempts to commit crimes within each category. Some merely provide that the penalty for attempt may be as great as for the completed crime. As to statutes dealing with attempts to commit particular crimes, the authorized punishment is usually lower than for the completed crime, but in some instances the same or even a higher punishment is possible.

*Id.* (footnotes omitted).

68. See *South Carolina v. Gathers*, 490 U.S. 805, 819 (1989) (O'Connor, J., dissenting) *overruled on other grounds* by *Payne v. Tennessee*, 501 U.S. 808 (1991). Justice O'Connor noted the narrowness of the distinction between a failed and a successful attempt:

In the death penalty context, no State authorizes infliction of the penalty for attempted murder, yet the criminal defendant who has attempted to kill another human being has the same mental state as the actual killer. Indeed, as Justice Scalia noted in dissent in *Booth*, the difference between murder and attempted murder may often hinge on a fortuity over which the defendant has no control at all. The only distinction is the harm to the community which results from the defendant's actions, and this distinction is deemed sufficient

miscarries deserves lesser punishment than one whose attempt succeeds.

For example, A and B both shoot their victims. A's victim receives ordinary medical care but dies nonetheless. B's victim happens to fall at the feet of a military surgeon with great experience in dealing with gunshot wounds, and survives as a result. A and B have acted with equal wickedness. On the face of their conduct they are equally dangerous (unless one supposes that B might have a continuing animus against his victim, who, having survived, makes B *more* of a risk than A).

The great weight of serious scholarship, including the highly influential ALI Model Penal Code, treats the attempt/success distinction as unjustifiable.<sup>69</sup> How then has it survived in positive law? FAE offers a powerful explanation. Legislators (and voters)<sup>70</sup> tend to attribute success or failure to the person, not the situation.<sup>71</sup>

---

to support a difference in punishment between a sentence of years and the ultimate penalty.”

*Id.* (internal citations omitted).

69. See, e.g., Larry Alexander, *Crime and Culpability*, 5 J. CONTEMP. LEGAL ISSUES 1, 3-20 (1994) (defending culpability *vis a vis* harm caused as determining blameworthiness); Kadish, *supra* note 66, at 680 (arguing that the harm doctrine is not “rationally supportable” meaning that “it is a doctrine that does not serve the crime preventive purposes of the criminal law, and is not redeemed by any defensible normative principle”); MODEL PENAL CODE § 5.05 (Proposed Official Draft 1962) (generally equating attempts with completed offenses in applicable penalties). The MPC, however, makes two concessions to the significance of harm actually caused. First, section 5.05(1) provides that an attempt to commit a capital crime is not itself capital, and section 5.05(2) provides that an inherently improbable attempt may, in the discretion of the court, be mitigated or excused altogether. *Id.*

Michael Moore, who himself defends the role of moral luck in current law, writes that since the promulgation of the MPC, “many criminal law theoreticians who have addressed the issue have concurred [with the ALI view], including Joel Feinberg, Stephen Schulhofer, Richard Parker, Michael Zimmerman, Hyman Gross, Andrew Ashworth, Judith Thomson, Steven Sverdlik, Lawrence Becker, James Gobert, Nicholas Rescher, and Sanford Kadish.” Michael S. Moore, *The Independent Moral Significance of Wrongdoing*, 5 J. CONTEMP. LEGAL ISSUES 237, 239-40 (1994) (footnotes omitted). Moore describes opposition to the harm doctrine as “the standard educated view.” *Id.* at 238. He presents an impressive resume of unsuccessful attempts to justify prevailing doctrine’s emphasis on actual results. See *id.* at 241-51. His case for grading culpable acts according to the harm caused rests primarily on intuitions about the degree of resentment and guilt felt in the wake of harmful and unarmful attempts or risk-takings. See *id.* at 263-71. Moore takes these intuitions as data points for the method of reflective equilibrium; the psychological research suggests that these same intuitions are due to cognitive predispositions entitled to no presumption of correctness.

70. Note again that the success/attempt distinction is one of the cases in which Robinson’s and Darley’s lay subjects rejected the MPC, aligning themselves with prevailing positive law. See *supra* note 29.

71. For the application of cognitive psychology to legislative decisions, see Jeffrey J. Rachlinski & Cynthia R. Farina, *Cognitive Psychology and Optimal Government Design*, 87 CORNELL L. REV. 549, 564-68 (2002). For a critique arguing that the psychological research does not yet permit normative prescriptions, see William N. Eskridge, Jr. & John Ferejohn,

Ignoring the fortuitous nature of the handy surgeon in B's case, the law determines whether to reward or punish B and A simply based on their moral luck.<sup>72</sup>

A second widely criticized doctrine that bases liability on harm actually caused is the felony murder rule. The rule holds all participants in a dangerous felony, such as robbery, burglary, or rape, guilty of murder when death is caused during the commission of the felony.<sup>73</sup> There are various limitations on the rule,<sup>74</sup> but these do not detract from its strict liability character.<sup>75</sup> The getaway driver, with no intent to kill, and even without awareness of any such intention on the part of his accomplices, is deemed guilty of murder when an accomplice kills (or, for that matter, is killed) inside the bank or the store. The rule remains an established part of positive law despite sustained and cogent criticism.<sup>76</sup> Indeed, the modern scholarship seems less interested in whether the rule should pass away than in why it hasn't.<sup>77</sup>

---

*Structuring Lawmaking to Reduce Cognitive Bias: A Critical View*, 87 CORNELL L. REV. 616 (2002). The attempt/success distinction seems like a strong candidate for attributing legislative choice to cognitive predisposition because the public choice incentives point in the other direction. Felony murder might be the product of law-and-order politics pure and simple, but the attempt/success distinction is not so easily explained.

72. Moore recognizes that there is some tension between his powerful critique of *all* prior defenses of the harm doctrine and his willingness to use the doctrine's hitherto unjustified acceptance as itself a justification. Moore, *supra* note 69, at 266-67. Coupled with his preferred method of justification, his critique of prior justifications offered for the harm-doctrine fits quite nicely with the thesis that the harm doctrine rests less on widely shared moral intuitions and more on widely shared cognitive tendencies.

73. See, e.g., LAFAVE, *supra* note 9, § 7.5, at 671-76.

74. *Id.* at 672-81.

75. Some commentators distinguish between a strict liability interpretation of the rule and an interpretation premised on substitute culpability or constructive malice. See, e.g., Nelson E. Roth & Scott E. Sundby, *The Felony-Murder Rule: A Doctrine at Constitutional Cross-roads*, 70 CORNELL L. REV. 446, 453-60 (1985) (explaining this distinction). An intent to rob, however, is neither the intent to kill nor, *ipse facto*, depraved indifference to human life. The effect of the rule, however worded in a given statute or jury instruction, is to hold the felons strictly liable as murderers when death results. See Note, *Felony-Murder: A Tort Law Reconceptualization*, 99 HARV. L. REV. 1918, 1918 (1986) ("A comparison with the principles of tort law will show that felony murder defendants receive considerably less protection from the law than defendants in wrongful death actions.").

76. Denying the attractiveness of the intent-based retributivists' challenge to the special rules governing killings during crime is no easy task. But even conceding the power of those arguments, they are indisputably being ignored. Courts and legislatures seem committed to this form of strict liability in criminal law. Kevin Cole, *Killings During Crime: Toward a Discriminating Theory of Strict Liability*, 28 AM. CRIM. L. REV. 73, 77 (1990).

77. Roth & Sundby, *supra* note 75, at 448 ("The felony-murder rule's continued vitality despite articulate criticisms of both its rationale and its results indicate [*sic*] that a policy analysis alone will not abate its use") (proposing constitutional attacks that likewise have not succeeded); James J. Tomkovicz, *The Endurance of the Felony-Murder Rule: A Study of the Forces that Shape Our Criminal Law*, 51 WASH. & LEE L. REV. 1429, 1431 (1994) ("My goal is

Part of the explanation for the survival of the felony murder rule is political. The rule helps prosecutors, who have great influence with legislators, and hurts felons, who have little.<sup>78</sup> The rule reduces the prosecution's burden of proof and encourages confessions from suspects foolish enough to believe that only the person who did the killing is liable for the murder.

FAE may also play a significant role in the rule's survival. Legislators and judges may very well focus on behavior, rather than situation, and therefore attribute death to the non-shooting accomplices rather than to the unexpected resistance of the victim, intervention of the police, accidental discharge of a weapon, or collision during a chase. Alternatively, they may reasonably believe that the electorate will make a similar judgment.<sup>79</sup>

The criminal law scholars have asked: "How is the felony-murder defendant different from any other robbery defendant?" The rational answer is that, absent some *mens rea* with respect to the killing, as distinct from the robbery, there is no difference. But FAE inclines the observer against the rational conclusion. The difference is

---

more modest: to understand how a rule of law that has been maligned so mercilessly for so long and that is putatively irreconcilable with basic premises of modern criminal jurisprudence has survived and promises to persist into the twenty-first century." (footnote omitted).

78. See Rudolph J. Gerber, *The Felony-Murder Rule: Conundrum Without Principle*, 31 ARIZ. ST. L.J. 763, 767 (1999) ("Notwithstanding trenchant criticism, the rule still operates throughout the United States because lawmakers lack the courage to amend or repeal it."); Tomkovicz, *supra* note 77, at 1460-65 (explaining power of law-and-order politics and heightened popular concern when human life is lost).

79. See Tomkovicz, *supra* note 77, at 1477.

These strains of popular thought concerning culpability are intertwined and inseparable. The reason for exploring them here is not to suggest that the public finds scholarly conceptions of fault, blame, and criminal punishment generally unacceptable as the basis for a system of criminal laws. The law on the whole incorporates and is consistent with the culpability principles predominant in scholarly thought. The point is that there are additional premises that are integral parts of the public psychology. Those premises help account for the contemporary retention of a doctrine descended from prior generations and compatible with modern American politics. They help explain the inefficacy of the scholarly community's accusations that felony-murder transgresses basic principles and the failure of numerous efforts to overthrow the rule.

*Id.* (footnote omitted).

The Darley & Robinson findings on felony murder suggest that the popular view may be closer to the academic view than Professor Tomkovicz supposed. Compare *id.*, with *supra* note 29. The Darley & Robinson subjects would only have imposed strict liability for manslaughter in the felony murder context. ROBINSON & DARLEY, *supra* note 29, at 178. Even this, however, is a clear departure from the culpability principle. Professor Tomkovicz took a somewhat agnostic view about the ultimate merits of the popular as opposed to the academic view. My point here is that research on FAE gives us some reason to regard opinion supporting the felony murder rule, whether popular or professional, as determined more by cognitive predisposition than by any plausible moral judgment.

that the felony murder defendant's robbery resulted in death, and it follows that the person, not the situation, must be to blame for it.

One key to any plausible defense of grading criminal liability according to harm caused is a strong proximate cause limit.<sup>80</sup> Absent such a limit, the lottery-like character of strict liability doctrines such as felony murder becomes more pronounced. FAE, together with the hindsight bias, however, suggests that proximate cause—notoriously difficult to define more precisely than as an appeal to intuition—will tend to be applied expansively rather than restrictively. The modern felony murder case law, for instance, includes appellate opinions upholding convictions on rather improbable causal sequences.

Courts have upheld felony murder convictions predicated on aggravated possession of a stolen vehicle, when, following a high-speed chase, the defendant abandoned the stolen car on foot, and was then pursued on foot by an officer who was run down by the vehicle of another;<sup>81</sup> predicated on complicity in a burglary when the principal's entry precipitated a heart attack by a resident;<sup>82</sup> and predicated on rape and assault when the elderly victim became depressed, lost appetite, and choked to death during assisted feeding in a nursing home a month after the attack.<sup>83</sup> These decisions necessarily follow trial verdicts favorable to the government, revealing that juries as well as judges agreed with the attribution of responsibility to the accused.<sup>84</sup> Reliance on proximate cause to limit harm-based liability would appear, in practice, to be misplaced.

---

80. See Alexander, *supra* note 69, at 13-17 (arguing that proximate cause reduces to a culpability-based limit on harm-caused liability, and that without such a limit, harm-caused liability would be too unfair even for its proponents).

81. *People v. McCarty*, 769 N.E.2d 985 (Ill. App. 2002), *vacated*, 785 N.E.2d 859 (Ill. 2003), *directed to reconsider in light of* *People v. Balk*, 784 N.E.2d 825 (Ill. 2003).

82. *People v. Howard*, 661 N.Y.S.2d 386, 386-88 (App. Div. 1997) (holding that the government needed only to prove participation in the burglary and the cause of death to establish felony murder liability).

83. *Brackett v. Peters*, 11 F.3d 78 (7th Cir. 1993).

84. Professor Finkel's study of mock jurors given felony murder scenarios suggests substantial willingness to nullify the felony murder rule. See FINKEL, *supra* note 3, at 164-69. The study design, however, gave each juror four defendants to judge, one the triggerman and the others accomplices. "Thus, although the subjects judged the triggerman harshly, they did not transfer their judgments to the accessories." *Id.* at 166. Given the scenario, the subjects could excuse the accomplices while still insisting on personal as opposed to situational causation. *Cf. Tison v. Arizona*, 481 U.S. 137, 139-41 (1987) (upholding capital sentences for felony murder convictions arising out of prison escape). In *Tison*, the father of the defendants had died of exposure in the desert following a shootout with police. *Id.* at 141. With the chief culprit removed from the scene, the search for personal rather than situational causes would turn to his surviving sons, even though neither had actually shot the victims. *Id.* In any event, Professor Finkel may be overstating the degree of community hostility to the felony murder doctrine. There is no pressure to abolish it and the reported cases include too many convictions to believe that widespread nullification is taking place.

A final doctrinal anomaly involving harm actually caused is the so-called *Pinkerton* doctrine. The Supreme Court has held that members of a criminal conspiracy are *ipse facto* liable as accomplices for the substantive offenses committed by coconspirators in furtherance of the conspiracy, so long as these substantive crimes were not unforeseeable.<sup>85</sup> The peculiar feature of the doctrine is that the defendant remains liable as an accomplice even after he has been arrested and jailed.<sup>86</sup> One can imagine hypotheticals in which this result makes sense, such as when the defendant hires a contract killer, is then arrested, and the killer follows through on the plot. In the ordinary *Pinkerton* case, however, liability attaches to crimes that were merely foreseeable, rather than actually foreseen, by the accused. This broad view of human agency is more consistent with FAE than with any rational justification. Reflecting the need to hold someone responsible for negative outcomes, the law attributes responsibility to the person it has in custody, rather than to those whose independent actions are the dominant causes of the new offenses.

#### IV. EXPLAINING THE DEFENSES

##### A. *Personal and Impersonal External Causes: Insanity, Necessity, and Intoxication Contrasted With Duress, and the Abuse Excuse*

The criminal law sometimes excuses the intentional or reckless causing of harm. Insanity,<sup>87</sup> necessity,<sup>88</sup> and duress<sup>89</sup> are well-known examples. The plea of duress or insanity admits that the defendant intended to and did cause harm. Responsibility for the defendant's conduct, however, lies outside his moral agency, with an identifiable

---

85. *Pinkerton v. United States*, 328 U.S. 640, 647-48 (1946).

86. *Id.* The *Pinkerton* doctrine's recent use in litigation has been to impose on conspirators the enhanced penalties provided by Congress for the use of firearms during crimes, so that any conspirator, even one not present at the time of the substantive offense, is held guilty of carrying and using a firearm during the offense. See, e.g., *United States v. Diaz*, 248 F.3d 1065, 1098-99 (11th Cir. 2001); *United States v. Bailey*, 235 F.3d 1069, 1074-75 (8th Cir. 2000).

87. See LAFAVE, *supra* note 9, §§ 4.1-4.7, at 321-91 (discussing insanity defense); 2 PAUL H. ROBINSON, *CRIMINAL LAW DEFENSES* § 173, at 280-313 (1984) (same).

88. See, e.g., LAFAVE, *supra* note 9, § 5.4, at 476-86 (discussing necessity defense); 2 ROBINSON, *supra* note 87, § 124, at 45-68 (same).

89. See, e.g., LAFAVE, *supra* note 9, § 5.3, at 467-76; 2 ROBINSON, *supra* note 87, § 177, at 347-72.

disease<sup>90</sup> or a third party's threat.<sup>91</sup> From a retributive point of view, however, the scope of these defenses is not consistently defined.

The different legal treatment of duress and necessity is illuminated by FAE. Duress requires an immediate threat, one that would overcome the resistance of an ordinary person, even if the defendant is not an ordinary person. A defense of duress is generally not available to counter a charge of murder.<sup>92</sup> The general choice-of-evils defense, which was not generally recognized by the common law courts,<sup>93</sup> is not universally recognized by American legislators,<sup>94</sup> and is grudgingly interpreted by the courts.<sup>95</sup> Duress, with its focus on the blameworthy threats of a third party,<sup>96</sup> has roots in the common law<sup>97</sup> and is pervasively recognized in contemporary positive law.<sup>98</sup> The more sympathetic treatment of duress may reflect a greater

---

90. See, e.g., SANFORD H. KADISH, *BLAME AND PUNISHMENT* 100 (1987) ("Seen in this way it is apparent why the excuse of legal insanity is fundamental. No blaming system would be coherent if it imposed blame without regard to moral agency. We may become angry with an object or an animal that thwarts us, but we can't blame it.").

91. The duress defense was once viewed as a justification based on choosing the lesser of two evils. See LAFAVE, *supra* note 9, § 5.3, at 473. The modern view is that "[t]he excusing condition in duress is the impairment of the actor's ability to control his conduct." 2 ROBINSON, *supra* note 87, § 177(b), at 351. The Model Penal Code makes the defense available when the defendant was faced with a threat of force "that a person of reasonable firmness in his situation would have been unable to resist." MODEL PENAL CODE § 2.09 (Proposed Official Draft 1962).

92. See KADISH, *supra* note 90, at 93-94.

Here too there is an apparent gap between excuses allowed by the law and the requirements of just blaming. . . . The common law and the law of most states excuse in only one situation of this kind—duress . . . and even then, under the common law and the law of most jurisdictions, the act is excused only when the threat is immediate and the crime is not a homicidal one.

*Id.* at 93. (footnote omitted).

93. *Regina v. Dudley*, 14 Q.B. 273 (1884) (rejecting necessity defense to murder charge arising from killing of human being for food by starving castaways). Early English authorities referred to necessity but probably comprehended self-defense under this rubric. See JEROME HALL, *GENERAL PRINCIPLES OF CRIMINAL LAW* 416-17 (2d ed. 1947).

94. See 2 ROBINSON, *supra* note 87, § 124(a), at 45 ("The lesser evils defense, sometimes called 'choice of evils' or 'necessity' or the general justification defense, is recognized in about one-half of American jurisdictions." (footnote omitted)).

95. See *United States v. Oakland Cannabis Buyers' Coop.*, 532 U.S. 483, 489-95 (2001) (concluding that legislature implicitly excluded necessity defense when adopting Controlled Substances Act); *United States v. Bailey*, 444 U.S. 394, 409-17 (1980) (recognizing necessity defense to charge of escape, but refusing to permit jury to consider either duress or necessity without showing that escape was only means possible of avoiding greater evils and that the accused promptly return to custody).

96. KADISH, *supra* note 90, at 93 (referring to excuses "based on reasonable deficiency of will" (footnote omitted)).

97. See 4 WILLIAM BLACKSTONE, *COMMENTARIES* \*30.

98. 2 ROBINSON, *supra* note 87, § 177(a), at 348 ("Nearly every American jurisdiction recognizes some form of duress excuse." (footnote omitted)).

willingness to attribute negative outcomes to human beings rather than to impersonal factors.

Even in the days when many jurisdictions accepted the ALI's definition of the insanity defense, by recognizing a volitional as well as a cognitive prong to the excuse,<sup>99</sup> the accused could not claim either prong without first establishing a recognized medical disease or defect.<sup>100</sup> Whatever the effect on behavior or cognition, clinically recognized "disorders" do not by themselves support an insanity plea.<sup>101</sup> The contemporary return to the *M'Naghten* test holds accountable even a defendant who acted in response to an irresistible impulse caused by a recognized species of mental illness.<sup>102</sup>

The Hinckley case and its aftermath provide an excellent example of person/situation attributions in the criminal justice system.<sup>103</sup> At the trial, the jurors found Hinckley not guilty by reason of insanity—a verdict FAE predicts jurors would be reluctant to reach. The defense, however, had a strong case; leaving the psychiatric technicalities to one side, Hinckley seemed to fit the popular understanding of crazy.<sup>104</sup> As for behavioral control, Hinckley pulled the trigger with the proverbial policeman at his elbow.<sup>105</sup>

---

99. See Model Penal Code § 4.01(1) (Proposed Official Draft 1962) ("A person is not responsible for criminal conduct if at the time of the conduct as a result of mental disease or defect he lacks substantial capacity to appreciate the criminality [wrongfulness] of his conduct or to conform his conduct to the requirements of law." (brackets in original)).

100. *Id.* at 98 ("The modern tests of legal insanity are varied and controversial, but they all rest on the view that the claim of incapacity to comply with the law—because of defects of understanding or self-control—is an excuse only if it is the result of a mental disease." (footnote omitted)). The definition of "mental disease or defect" has been fraught with uncertainty. See LAFAVE, *supra* note 9, § 4.3(c), at 347-48; 2 ROBINSON, *supra* note 87, § 173(b), at 286-90.

101. Cf. P. Moran, *The Epidemiology of Antisocial Personality Disorder*, 34 SOC. PSYCHIATRY & PSYCHIATRIC EPIDEMIOLOGY 231, 234 (1999) ("Over the past decade, however, there have been a number of well-conducted surveys . . . These studies show that antisocial personality disorder is extremely common in prisons with prevalence rates as high as 40-60% among the male sentenced population.").

102. The *M'Naghten* test in pristine form excludes any consideration of behavioral control. See, e.g., LAFAVE, *supra* note 9, § 4.2, at 340 ("After *M'Naghten*, English trial judges instructed juries only in terms of the right-wrong test and 'irresistible impulse' was expressly rejected as 'a most dangerous doctrine' in 1863." (footnote omitted)).

103. For the evidence and the verdict, see generally LOW ET AL., *supra* note 2.

104. The basic irrationality of assassinating the President to win the love of a movie star is widely recognized. As defense counsel put it in closing argument:

In his own mind, defendant had two compelling reasons to do what he did. To terminate his own existence and to accomplish his ideal union with Jodie Foster, whether it be in this world or the next . . . [In] a classic understatement, Dr. Dietz stated that these were not the reasonable acts of a completely rational individual.

LOW ET AL., *supra* note 2, at 103 (ellipsis in original) (citing closing argument of defense counsel Vincent Fuller). A survey conducted shortly after the trial found that 73.3% of respondents thought that Hinckley was guilty, but that only 65.7% thought he was insane. 59.5% thought he

The decision, however, provoked a political reaction that sharply curtailed the insanity defense in both the federal system<sup>106</sup> and most of the states.<sup>107</sup> Legislators, following popular opinion, viewed the acquittal as a miscarriage of justice.<sup>108</sup> Only by attributing behavior to the individual, or construing mental illness as personal rather than situational, can this conclusion make any sense.<sup>109</sup> While social control interests are at stake in criminal justice, they play only a decidedly minor role in the insanity context, where acquittal leads to civil commitment<sup>110</sup> (where Hinckley remains to this day).<sup>111</sup> The risks

---

should be both punished and treated psychiatrically, while 14.1% said he should be treated but not punished. The respondents, however, appeared not to understand either the legal definition of insanity or to realize that Hinckley's commitment would be indefinite. Valerie P. Hans & Dan Slater, *John Hinckley, Jr. and the Insanity Defense: The Public's Verdict*, 47 PUB. OP. Q. 202, 206 (1982). The numbers are consistent with FAE; about ten percent of the respondents would hold Hinckley responsible even though they characterized him as insane, and a majority of 59.6% would hold him responsible even though he needed psychiatric therapy.

105. See LOW ET AL., *supra* note 2, at 84 ("When you look at the videotape you see a Secret Service agent walks right up to Mr. Hinckley and turns to the left and it is at that point after the agent turns his back away that Mr. Hinckley pulls out the gun and fires." (citing closing argument of government attorney Roger Adelman)).

106. See Insanity Defense Reform Act, 18 U.S.C. §17 (2000) (reinstating *M'Naghten* test).

107. Sanford H. Kadish, *Fifty Years of Criminal Law: An Opinionated Review*, 87 CAL. L. REV. 943, 960-61 (1999).

Enough states reverted to *M'Naghten* that it displaced the Code's formulation as the states' majority rule, and Congress saw fit to dump the Model Penal Code test for all federal prosecutions, substituting another version of *M'Naghten*. But that was not all. The move to restrict the insanity defense included procedural changes in many jurisdictions, such as shifting the burden of proof of insanity to the defendant (a reform now adopted by most jurisdictions), making it harder for insanity acquittees to obtain their release, introducing a new "guilty but mentally ill" verdict, and even, in the case of a few states, abolishing the defense entirely.

*Id.* (footnotes omitted).

108. See, e.g., GEORGE E. DIX & M. MICHAEL SHARLOT, CRIMINAL LAW: CASES AND MATERIALS 855-59 (5th ed. 1996); Michael L. Perlin, *Unpacking the Myths: The Symbolism Mythology of Insanity Defense Jurisprudence*, 40 CASE W. RES. L. REV. 599, 609 (1990).

109. Kadish, *supra* note 107, at 960.

The main target of the attack was the volitional prong of the Model Penal Code's formulation. Apparently it's easier to accept that a person's mental impairment may disable him from knowing what he did than that it prevents him from willing the criminal action he performed. The latter cuts too deeply into how we conceive of human behavior.

*Id.*

110. See, e.g., LAFAYE, *supra* note 9, § 4.1, at 323.

The insanity defense is quite different from other defenses in that the result, if it is successfully interposed, is not acquittal and outright release of the accused but rather a special form of verdict or finding ("not guilty by reason of insanity") which is usually followed by commitment of the defendant to a mental institution. Thus, its purpose is usually said to be that of separating from the criminal justice system those who should only be subjected to a medical-custodial disposition.

that the civil system will restrain less effectively than the criminal system, or the risk that potential criminals, planning to fake the defense, would rather be committed than imprisoned, seem quite remote.<sup>112</sup> The best explanation for the return to *M'Naghten* may very well be FAE—public opinion and legislators following it attribute crime to the person rather than the disease.

Strikingly, at the same time that the law has held the mentally ill responsible for their behavior, it has also moved in the direction of adopting sexual predator laws that authorize indefinite civil commitment based on the propensity to commit future offenses.<sup>113</sup> Indeed, the Supreme Court now holds that mental abnormality producing “serious difficulty in controlling behavior” permits indefinite civil commitment.<sup>114</sup> If mental disorders are thought powerful enough to predict future behavior with sufficient confidence to commit the legally innocent to institutions as a preventive measure, one might suppose that similar conditions might furnish an excuse for crimes actually committed.<sup>115</sup>

---

*Id.*

111. Cristy Gutowski, *Why Lemak Has Long Odds With Rare Insanity Defense*, CHI. DAILY HERALD, Nov. 13, 2001, at 1 (“Some argue John Hinckley, found not guilty by reason of insanity for his attempt to assassinate President Ronald Reagan in 1981, might be out of prison today had he been convicted. Instead, Hinckley remains confined to St. Elizabeth’s Hospital in Washington, D.C.”), available at 2001 WL 30249936.

112. The Supreme Court has held that the civil commitment following an insanity acquittal can exceed the maximum time that might have been imposed following a conviction. *Jones v. United States*, 463 U.S. 354, 368-69 (1983). Pleading the defense thus runs the risk of indefinite detention. See also LaFave, *supra* note 9, § 4.6, at 390-91 (noting that “in most jurisdictions the burden of proof is on the patient, in habeas corpus or other statutory proceedings for release, to persuade the court that he meets the statutory criteria and is thus entitled to release.”). While the burden on the patient is particularly difficult when the hospital staff do not support his request, even the support of the hospital professionals often does not persuade courts that doubt the ability of those professionals to evaluate dangerousness. *Id.* at 391.

113. See, e.g., Paul H. Robinson, *Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114 HARV. L. REV. 1429, 1430 n.6 (2001) (At the time *Kansas v. Hendricks*, 521 U.S. 346 (1997), was argued, “six states had such [sexual predator commitment] statutes—the other five [besides Kansas] being Arizona, California, Minnesota, Washington, and Wisconsin—and thirty-eight states, including New Jersey and New York, filed amicus briefs successfully urging the Justices to uphold the law.”).

114. *Kansas v. Crane*, 534 U.S. 407, 413 (2002).

115. Professor Morse exposes the contradiction succinctly: “[a]n agent responsible enough to warrant criminal punishment is sufficiently responsible to avoid preventive detention.” Morse, *supra* note 18, at 259. The justices of the Supreme Court are probably fully aware of the inconsistency here; the *Crane* test of eligibility for commitment seems indistinguishable from the ALI’s insanity defense, but Justice Breyer was careful to say “serious difficulty controlling behavior” rather than “lacks substantial capacity.” *Crane*, 534 U.S. at 413.

Many factors contribute to the political prevalence of this anomaly. Bona fide social control interests,<sup>116</sup> magnified by law-and-order politics,<sup>117</sup> certainly play their role. FAE, however, may very well play a role as well. The distinction between narrow definitions of the insanity defense and the willingness to confine based on propensity bears a certain resemblance to the quiz master experiment. Intellectually, legislators and voters know that criminal behavior is powerfully induced, and they show that awareness by adopting sexual predator laws. At another level, however, they cannot get over attributing to the defendant responsibility for the very acts the sexual predator laws suppose to be beyond his control.

The so-called "abuse excuse," the plea that a homicide victim's mistreatment of the defendant excuses (or even justifies) the homicide, is not a single defense but a ganglia of doctrinal innovations. The most important of these doctrines broaden the grounds of self-defense by relaxing the imminence requirement,<sup>118</sup> and admitting social framework evidence in the form of expert opinions (syndrome

---

116. "Both [James Allen Egelhoff and Leroy Hendricks] committed devilish deeds and both appear to be very dangerous people. Few people will shed a tear for them if the Court's decisions weigh public safety more heavily than culpability." Morse, *supra* note 18, at 253.

117. See David J. Gottlieb, *Preventive Detention of Sex Offenders*, 50 U. KAN. L. REV. 1031, 1047-48 (2002).

Like the Washington SVP Act, which Kansas adopted practically verbatim, the Kansas SVP Act was the product of intense public outcry after the commission of a terrible crime by a former prisoner who had been released from his criminal sentence. In each state, task forces were formed, with the purpose of designing more severe sentences for individuals accused of sexual offenses. In Kansas, the primary expression of this concern was a toughening of criminal penalties. Thus, the State added to its determinate sentencing law an enhancement for individuals designated as sexually violent predators, increased sentences for violent offenders, and enacted a death penalty bill. But the task force struggled with the question of how to continue confinement of individuals who were entitled by law to their freedom at the end of their sentence. The task force wished to increase the sentence of these individuals but was unable to do so by criminal sanction because of the constitutional prohibition against *ex post facto* laws and double jeopardy. The task force therefore turned to possible civil remedies. The outcome of this concern was the SVP Act, a statute which promised the continued confinement of "dangerous sex offenders . . . past their scheduled prison sentence." Although confinement under the Act was designated as civil, it was troubling that its confinement was, in essence, limited to those already convicted of sex offenses. It was also troubling that the "civil" treatment did not commence until the conclusion of the inmate's criminal sentence.

*Id.* (footnotes omitted).

118. See, e.g., *Developments in the Law: Legal Responses to Domestic Violence*, 106 HARV. L. REV. 1574, 1582 (1993) (noting that even jurisdictions with an imminence standard will sometimes consider the known violence of an assailant or testimony regarding the battered woman syndrome).

evidence).<sup>119</sup> Tactically, the defense aims to cast blame on the deceased, without invoking the plea of involuntary manslaughter.

The "abuse excuse" has provoked heated debate. Distinguished critics, including Alan Dershowitz and James Q. Wilson, argue that the "abuse excuse" implicitly rejects, and thereby threatens more generally, the criminal law's fundamental premise of free will.<sup>120</sup> What this criticism neglects is that expansive notions of excuse have thrived only when the defendant can cast blame on another human being. Paradoxically, the defendant can persuade judges, jurors and legislators to blame the victim only on the assumption that the victim had free will.

When the defendant points to situational factors *other than wrongdoing by another human being*, the law's understanding turns into condemnation. Intoxication, even when convincingly characterized as a disease, is not recognized as an excuse, even to such petty offenses as public drunkenness.<sup>121</sup> Just as the law condemns the completed offense more severely than its fortuitous failure, so the law equates the choice to get drunk (even when that choice was induced by pathology) with the choice to kill, rape, or rob.<sup>122</sup> Only this sort of moral equation can justify treating one who kills because intoxicated with one who kills because greedy or sadistic. Yet drunkenness, by

---

119. See, e.g., CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE § 7.22, at 983 (2d ed. 1999) ("Many cases approve evidence of battered woman syndrome offered to support claims of self-defense where it sheds some light on defendant's subjective fear and its nature and reasonableness."). Professor Mosteller has persuasively traced the receptivity to social framework evidence to a political determination that the balance of advantage should be shifted in certain categories of cases. See Robert P. Mosteller, *Syndromes and Politics in Criminal Trials and Evidence Law*, 46 DUKE L.J. 461, 478-91 (1996). If that is so, then it is substantially accurate to treat the reception of social framework evidence as an indirect way of modifying the substantive criminal law.

120. See generally DERSHOWITZ, *supra* note 16; JAMES Q. WILSON, MORAL JUDGMENT (1997).

121. See 2 ROBINSON, *supra* note 87, § 176, at 337-38 (leaving aside situations in which intoxication may be inconsistent with an element of the charged crime, a few jurisdictions provide no defense of intoxication, most provide a defense limited to involuntary intoxication, and a few permit intoxication that has the same cognitive or behavioral effects as required to show legal insanity). The use of intoxication to negate an element of the charge is limited by a variety of doctrines. See 1 PAUL H. ROBINSON, CRIMINAL DEFENSES § 65(a), at 292-93 (1984). In some places intoxication can negate specific but not general intent; in others it can negate intent but not recklessness; still others forbid consideration of intoxication altogether even when intoxication is alleged to be inconsistent with specific intent. *Id.* The Supreme Court upheld this practice in *Montana v. Egelhoff*, 518 U.S. 37, 41-43 (1996).

122. See 2 ROBINSON, *supra* note 87, § 162(b), at 250 (1984) ("[T]he imputation of recklessness is objectionable because even if the actor is reckless, or even purposeful, as to *getting intoxicated*, it does not follow that he is in fact reckless as to *causing the death of a pedestrian*. The notion that a person risks all manner of resulting harm when he voluntarily becomes intoxicated is common, but it is obviously not necessarily correct." (footnote omitted)).

itself, is not a crime, a fact that comports at best uneasily with any moral equation of drinking and killing.

The “abuse excuse,” then, does not reflect any lack of respect for the law’s assumption of free will. Rather, the refraction of the free will postulate through cognitive processes that exaggerate, rather than trivialize, personal responsibility goes a long way toward explaining the success of the “abuse excuse” relative to the success of claims founded on situational, but nonhuman causative factors. A rational assessment of blameworthiness might very well approve of the “abuse excuse.” It is hard to imagine, however, any rational assessment of blameworthiness that approved of the “abuse excuse” that would not also require legal excuses for intoxication and necessity far broader than those available under current law.

*B. Causing the Condition of One’s Own Defense: Forfeiture Contrasted With Culpability Analysis*

Frequently, the defendant asserting either excuse or justification may have brought himself into the situation in which he was pressured into harmful conduct. Generally speaking, current law holds this against the accused, typically by prohibiting the defendant from raising the defense *at all* if he was at fault, even if only to the degree of negligence, in creating the excusing or justifying condition.<sup>123</sup> Some jurisdictions go further and do not require any fault with regards to creating the condition at all when the accused could have avoided the excusing situation.<sup>124</sup>

As Professor Robinson has argued, the logical approach from the standpoint of the culpability principle would be to grade the actor’s liability according to the degree of fault shown in the creation of the excusing condition.<sup>125</sup> Thus, one who negligently commences an affray or exposes himself to the risk of duress might be guilty of a crime of negligence, but not of an intentional or reckless one. The law, however, appears to follow the forfeiture-by-wrongdoing model at least as often as it follows culpability analysis.

F AE may have something to do with the forfeiture model’s popularity. Judges and legislators may be attributing the ultimate harm to the actor’s original choice and ignoring situational variables that develop downstream. For purposes of illustration, consider the case of an attempted murder or aggravated assault defendant who

---

123. See Paul H. Robinson, *Causing the Conditions of One’s Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1, 8-14 (1985).

124. See *id.* at 2-8.

125. See *id.* at 27.

started the fight with non-deadly force, and whose adversary raises the ante by drawing a knife. The defendant draws a gun and wounds the victim. Under the forfeiture model, the defendant may not raise self-defense at all. Yet had the deceased prevailed in the struggle, he too would be guilty of attempted murder or aggravated assault, for the escalation from reasonable to deadly force would forfeit *his* self-defense claim. In the eyes of the law, whoever survives bears the blame more properly apportioned between the combatants.

### C. *Mitigating Circumstances in Capital Cases*

One final topic merits discussion under the general rubric of the excuses. Generally, the Supreme Court's death penalty jurisprudence holds that murderers may be executed so long as the sentencing tribunal has discretion to withhold the death penalty and so long as that discretion is guided by an individualized assessment of the aggravating and mitigating circumstances in each case.<sup>126</sup> Given that discretion to show mercy is also discretion to show none, there is considerable tension between the capital defendant's right to discretionary mercy and his right to have sentencing discretion exercised within the constraints of aggravating and mitigating factors set by statute or court decision.<sup>127</sup> FAE, however, suggests a perhaps more intractable problem with the Court's project of guiding discretion in capital cases.

FAE implies that capital sentencers will tend to underestimate situational factors that might mitigate a particular murder.<sup>128</sup> For

---

126. See, e.g., *Lockett v. Ohio*, 438 U.S. 586 (1978) (holding that a sentencer in capital cases must be allowed to consider mitigating factors offered by the defense even though factor is not recognized in applicable, facially valid statute); *Gregg v. Georgia*, 428 U.S. 153 (1976) (upholding constitutionality of guided discretion death penalty statute); *Woodson v. North Carolina*, 428 U.S. 280 (1976) (invalidating statute that made death penalty mandatory in defined categories of cases).

127. E.g., Carol S. Steiker & Jordan M. Steiker, *Sober Second Thoughts: Reflections on Two Decades of Constitutional Regulation of Capital Punishment*, 109 HARV. L. REV. 355, 382 (1995) ("This dilemma between *Gregg's* seeming insistence on channeling and *Woodson's* seeming insistence on uncircumscribed consideration of mitigating evidence constitutes the central dilemma in post-*Furman* capital punishment law.").

128. See Craig Haney, *Violence and the Capital Jury: Mechanisms of Moral Disengagement and the Impulse to Condemn to Death*, 49 STAN. L. REV. 1447, 1459 (1997).

Intensifying the effect of this perceived disparity in ability and the actual disparity in resources is the fact that the prosecution's implicit and overarching "theory" of the typical capital case generally comports with the jurors' stereotypical beliefs about crime and punishment. The notion that a defendant's crime stems entirely from his evil makeup and that he therefore deserves to be judged and punished exclusively on the basis of his presumably free, morally blameworthy choices is rooted in a longstanding cultural ethos

instance, Robert Alton Harris committed a horrific double murder, but only after a lifetime of brutalization and abuse that most of us simply cannot conceive.<sup>129</sup> The jury, in effect, attributed the crime to personality rather than situation. So too with the juries who have sentenced mentally retarded<sup>130</sup> or extremely youthful<sup>131</sup> defendants to death for murders that do not seem any more evil than murder always is.

The Court's basic aim in the death penalty cases seems to have become forcing the states to sift the population of convicted murderers and reserve the death penalty for the worst of the worst. A rational approach to this goal would be to limit the absolute number of death sentences any jurisdiction could impose, and leave it to the political authorities to prioritize by identifying those murderers who have earned the community's most intense condemnation. Such an approach would force decision makers to consider impersonal mitigating circumstances because the personal element is present in every case. Leaving the decision to juries in individual cases would be an odd way to pursue comparative reprehensibility determinations, even if juries rationally assessed the weight of extenuating situational factors. FAE implies that juries are not likely to make any such fully rational assessment. Whether one agrees with Justice Blackmun, who

---

that capital jurors (like most citizens) have been conditioned to accept uncritically. Add to this the well-documented tendency of most people to commit what psychologists have termed the "fundamental attribution error"—providing causal explanations for the behavior of others in largely dispositional or personal as opposed to situational or contextual terms. As a result, the typical juror's preexisting framework for understanding behavior is highly compatible with the basic terms of the typical prosecutorial narrative.

*Id.* (footnotes omitted).

129. David Dolinko, *Three Mistakes of Retributivism*, 39 *UCLA L. REV.* 1623, 1648-49 (1993).

What, for example, of the formative experiences of Robert Alton Harris, who murdered two San Diego teenagers in 1978 when he stole their car for use in a bank robbery, and finished their fast food hamburgers about fifteen minutes after the killings? Harris was the son of alcoholic parents, born prematurely after his drunken father kicked the pregnant Mrs. Harris in a jealous frenzy. This father, twice convicted of sexually molesting his daughters, beat his children frequently; Harris's mother—also an alcoholic and arrested several times—resented and abused him; he had a learning disability and a speech problem and, according to evidence adduced after his trial and death sentence, possible brain damage from fetal alcohol syndrome.

*Id.* (footnotes omitted).

130. *See, e.g.*, *Atkins v. Virginia*, 536 U.S. 304, 321 (2002) (holding execution of mentally-retarded offender unconstitutional). *Atkins* has put a stop to the practice, but twelve states executed thirty-five allegedly mentally retarded offenders during the period 1984-2000. *Id.* at 346-47 (citation omitted).

131. *See, e.g.*, *Stanford v. Kentucky*, 492 U.S. 361 (1989) (upholding death sentences of offenders who were, respectively, sixteen and seventeen at the time of the murders).

thought that the failure of the guided discretion regime called for abolition of the death penalty,<sup>132</sup> or with Justice Scalia, who thinks that the failure of the regime calls for a return to the days of unguided discretion,<sup>133</sup> FAE adds yet another reason to be skeptical of the regime.

## V. NORMATIVE IMPLICATIONS

### A. *Reducing Irrationality*

A review of doctrinal anomalies suggests that FAE operates throughout the criminal justice system. Before any given defendant can be punished, legislators must authorize the penalty, and judges and juries must agree to impose it. The opportunity to reduce the influence of cognitive predispositions is coextensively pervasive.

#### 1. Reconsidering the Significance of Resulting Harm

At the legislative level, three general directions deserve to be explored. First, the various doctrines that vary liability according to fortuitous results should be reconsidered. Critics have long belabored them, and no powerful defense has yet appeared. Their persistence is powerful evidence of FAE's role in criminal justice. Indeed, so strong is the attachment to these doctrines that I doubt that there is any practical chance of reform. For whatever chance it might provide, the harm doctrines should become widely understood as more consistent with FAE than with any rational principle.

#### 2. Distinguishing Situational, Third-Party, and Blame-the-Victim Defenses

Second, with respect to the excuses, legislators (and judges to the extent the judges are called upon to determine the scope of substantive defenses<sup>134</sup>) should distinguish situational excuses, third-

---

132. *Callins v. Collins*, 510 U.S. 1141, 1149 (1994) (Blackmun, J., dissenting from the denial of certiorari) (arguing that incompatible requirements of consistency across cases and fairness to individuals called for abolition of capital punishment).

133. *Id.* at 1141-42 (Scalia, J., concurring in denial of certiorari) (arguing that incompatible requirements of limited discretion and plenary consideration of mitigation called for end to guided-discretion regime).

134. In federal cases, the courts define affirmative defenses, subject to modification by Congress. The underlying theory is that although Congress has not adopted definitions of most defenses, Congress could not have intended liability in genuine cases of self-defense, entrapment, or whatnot. See, e.g., *Sorrells v. United States*, 287 U.S. 435, 447-49 (1932). Thus Congress,

party excuses, and blame-the-victim excuses. Insanity, intoxication, and necessity qualify as situational, because in these cases there is no human agent to whom responsibility can be assigned. Duress and entrapment fall into the second category; a human agent is blamed by the defense, but that agent is not the victim of the defendant's wrongdoing. When the defense attributes the crime to abuse inflicted on the accused by someone other than the victim of the charged offense, a similar classification is in order. Provocation in homicide cases, some claims of consent in rape cases, and the most common variations of the "abuse excuse" fall into the third category.

This classification does not have any rigorous doctrinal implications. There is plenty of perfectly rational normative disagreement about the exculpatory significance of the defendant's excuse in each instance. What should be recognized is that the willingness of observers to accept the defendant's claim will be weakest in situational cases and strongest in blame-the-victim cases. In particular, legislators should be more cautious of unjustified jury acquittals in cases with blame-the-victim defenses than in those with situational or third-party excuses. More specifically, the circumstances justifying an instruction on voluntary manslaughter and self-defense in homicide cases should be defined in light of the risk that jurors will attribute an exaggerated degree of causal responsibility to the deceased.

### 3. Rules Versus Standards

The third and final direction for legislative reforms is to regard with skepticism legal standards that appeal to intuitive notions of justice. Again, no specific doctrinal reform automatically follows. The "reasonable person" sets the standard for so many of the excuses because able lawyers simply could not articulate categorical rules that accurately captured the moral content of the excuse. The drafters of the Model Penal Code resorted to that formulation in many places.<sup>135</sup>

---

through initial silence and subsequent acquiescence, delegated the definition of defenses to the federal courts, subject to legislative modification of the sort that followed the *Hinckley* case. See *supra* notes 103-106 and accompanying text.

135. See, e.g., MODEL PENAL CODE § 2.13(1)(b) (Proposed Official Draft 1962) (entrapment defense available when government agents employ "methods of persuasion or inducement which create a substantial risk that such an offense will be committed by persons other than those who are ready to commit it"); § 2.09(1) (duress defense available when defendant acts because threatened with unlawful force against himself or another "that a person of reasonable firmness in his situation would have been unable to resist"); § 2.02(d) (actor is negligent when he disregards a risk "of such a nature and degree that the actor's failure to perceive it . . . involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation").

They also defined proximate cause in terms of events so improbable that punishment of the defendant for the result would simply be unjust.<sup>136</sup> Such formulations invite the fundamental attribution error, but it by no means follows that rigid categories would on the whole be any better.

Experience with the provocation defense is illustrative of the difficulties with both rigid limits on defenses and broad standards open to jury discretion. The common law defense's limited frame of time and cognizable provocations are largely arbitrary.<sup>137</sup> Under the more general and more normative rubric of the MPC, however, juries have been invited to blame the victim for her own death on account of such affronts as threatening or attempting to leave a dysfunctional relationship.<sup>138</sup> In choosing between rules and standards, legislatures defining crimes should account for the risk that standards will be applied by observers subject to FAE.

When more than one hallmark of the risk of cognitive error is in play, legislatures should be most skeptical. For example, legislators should be quite suspicious about relying on standards rather than rules to govern excuses of the blame-the-victim variety, as both the type of excuse and the amount of discretion given to juries in such a situation increase the influence of FAE. In an odd rearrangement of familiar political positions in homicide cases, this approach would cast doubt on both the provocation defense and on relaxing the imminent threat and duty-to-retreat rules in self-defense cases.

#### 4. The Trial Level

At the trial level, judges and lawyers should understand FAE and account for it in their choices. Again, however, there do not seem to be any simple procedural reforms with a great deal of promise. In the first place, much of FAE's influence operates at the legislative stage. So long as statutes and appellate opinions set up norms premised on exaggerated assessments of personal responsibility, the trial process can do little if anything to temper such doctrines as felony murder.

---

136. See § 2.03(2)(b) (conduct is legal cause of result when result is of same type as that foreseen and "is not too remote or accidental in its occurrence to have a [just] bearing on the actor's liability or on the gravity of his offense" (brackets in original)).

137. See, e.g., Norman J. Finkel & Jennifer L. Grocup, *When Mistakes Happen: Commonsense Rules of Culpability*, 3 PSYCHOL. PUB. POL'Y & L. 65, 74-77 (1997).

138. See Victoria Nourse, *Passion's Progress: Modern Law Reform and the Provocation Defense*, 106 YALE L.J. 1331, 1339-42 (1997).

The psychological literature includes some studies on debiasing strategies—efforts to reduce or eliminate cognitive irrationalities—that could possibly have applications in the context of FAE. Both collective decision procedures<sup>139</sup> and accountability—the knowledge that the decision will have to be defended in front of others<sup>140</sup>—have shown some promise as debiasing strategies. There are good reasons, however, to doubt that such debiasing strategies can succeed completely.

Currently, juries are collective bodies, and jurors know now that they will need to give reasons for their views during deliberations. The empirical evidence on jury use of prior crimes evidence nonetheless strongly suggests the persistent operation of FAE.<sup>141</sup> Similarly, legislatures are collective and deliberative bodies. We have seen, however, that many legislative decisions are more consonant with FAE than with rational policy judgments.<sup>142</sup>

Direct attempts to deal with FAE in the courtroom have serious drawbacks and limited potential effectiveness. Expert

---

139. See Dagmar Stahlberg et al., *We Knew It All Along: Hindsight Bias in Groups*, 63 *ORG. BEH. & HUM. DECISION PROCESSES* 46, 48 (1995) (“[R]esearch has shown that these cognitive biases [FAE, consensus underutilization, and base-rate fallacy] are attenuated in groups compared to individual judgment conditions.” (citations omitted)).

140. Phillip E. Tetlock, *Accountability: A Social Check on the Fundamental Attribution Error*, 48 *SOC. PSYCH. Q.* 227, 231 (1985) (noting that when observers knew they would need to justify their attributions of essay-writers’ attitudes, fundamental attribution error effect disappeared).

141. See FEIGENSON, *supra* note 3, at 61-62.

In criminal cases, evidence that the defendant has been convicted of other crimes in the past is sometimes admitted for the limited purpose of impeaching the defendant’s credibility as a witness. The admission is accompanied by an instruction that jurors not consider the prior convictions as evidence of the defendant’s guilt in the case at bar. Much experimental research confirms that mock jurors are unable to adhere to this limiting instruction. Those who sit on mock trials in which the priors are admitted with the cautionary instruction convict at a higher rate than those who do not learn about the priors. While not excluding other explanations, researchers frequently speculate that the fundamental attribution error may account for this effect: The priors lead jurors to conclude that the defendant is the sort of guy who commits crimes like this.

*Id.* (citations and footnote omitted).

In some of this empirical research, the jurors deliberated in groups. See Valerie P. Hans & Anthony N. Doob, *Section 12 of the Canada Evidence Act and the Deliberations of Simulated Juries*, 18 *CRIM. L.Q.* 235, 243 (1975-1976) (proof of prior conviction moved individual conviction rate from 40% to 45%, but moved group deliberation conviction rate from 0 to 40%). Moreover, the same phenomena was observed in the classic Kalven and Zeisel research, which was conducted on actual juries in real cases. See HARRY KALVEN, JR. & HANS ZEISEL, *THE AMERICAN JURY* 146-48 (1966) (conviction rate was 27% higher in cases in which jurors learned of defendant’s prior convictions).

142. See *supra* notes 66-122 and accompanying text.

testimony to alert jurors to the effects of FAE would be expensive, time consuming, and distracting. Given the reluctance of courts to order compensation of experts for indigent defendants,<sup>143</sup> the expense of expert testimony will cause it to leave behind many of the parties who might need it most. Expert testimony, although helpful, also apparently falls considerably short of guiding jurors to rational assessments of the reliability of eyewitness identifications, seemingly the closest analogy to the context of FAE.<sup>144</sup> Jury instructions (or pre-instruction education)<sup>145</sup> about cognitive tendencies might do no harm, but given the dubious effects of instructions about the law, there seems no reason to hope that they might do a great deal of good.<sup>146</sup> The empirical evidence regarding instructions about identification evidence<sup>147</sup> and prior convictions<sup>148</sup> is not encouraging.

The burden of proof might be assigned as a counterweight. The major role of the burden of proof, however, is with respect to the basic, disputed historical facts. Shifting the burden with respect to facts, because of doubts about the normative spin the jury may give those facts, runs the risk that many good claims of excuse will be rejected. Given the consensus that false positives are much worse than false negatives in criminal cases, increasing the burden of proof for the

---

143. See David Harris, *The Constitution and Truth Seeking: A New Theory on Expert Services for Indigent Defendants*, 83 J. CRIM. L. & CRIMINOL. 469, 486-87 (1992) ("Since lower courts have interpreted *Ake* [*v. Oklahoma*, 470 U.S. 68 (1985)] to mean that the defense receives assistance only when the accused's case will fail without it, the basic tools standard does very little for most indigent defendants.").

144. See BRIAN L. CUTLER & STEVEN D. PENROD, *MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY, AND THE LAW* 213-24 (1995).

145. Professor Haney envisions some sort of pre-instruction education of jurors regarding potential biases, prominently including FAE. See Craig Haney, *Making Law Modern: Toward a Contextual Model of Justice*, 8 PSYCHOL. PUB. POL'Y & L. 3, 30 (2002).

146. Professor Feigenson, for instance, doubts that jury instructions, even instructions drafted more clearly than many current ones, would be an effective antidote to cognitive tendencies. See Neal R. Feigenson, *The Rhetoric of Torts: How Advocates Help Jurors Think About Causation, Reasonableness, and Responsibility*, 47 HASTINGS L.J. 61, 163 n.334 (concluding that standard reasonable-person negligence instruction invites FAE, and that rewording instruction to focus on unreasonable risks rather than unreasonable persons might help, but "given jurors' general inability to understand instructions such changes might not significantly affect jurors' thinking"). For a review of studies casting doubt on jurors' comprehension of instructions, see *id.* at 66 n.9.

147. See, e.g., Steven Penrod & Brian Cutler, *Witness Confidence and Witness Accuracy: Assessing their Forensic Relation*, 1 PSYCHOL. PUB. POL'Y & LAW 817, 834 (1995) ("In summary, the experiments reviewed above provide little evidence that judges' instructions concerning the reliability of eyewitness identification improve juror decision making.").

148. See, e.g., Edith Greene & Mary Dodge, *The Influence of Prior Record Evidence on Juror Decision Making*, 19 LAW & HUM. BEHAV. 67, 76 (1995) ("Like many other researchers, we found that limiting instructions had little effect on jurors' use of this evidence. Verdicts from mock jurors with instructions about the restricted use of this information were not different from verdicts of jurors without such instruction.").

defense seems like a poor strategy for coping with cognitive predispositions.

Good lawyers, of course, have grasped the basic tactical implications of FAE without any need for the psychological literature. The best defense is a good offense, as Alan Dershowitz shows in print and as he and other able advocates have shown in practice.<sup>149</sup> Perhaps all that can be said from the standpoint of the interests of justice is that the trial judge should recognize the psychological dynamic at work, and, in borderline rulings on evidence and directed verdicts, incline against the party inviting the jury to blame the person rather than the situation.

*B. Social Cognition and Theories of Punishment: Is There a "Fundamental Retribution Error"?*

For readers expecting an elegant policy response to the influence of FAE in criminal justice, the preceding section may come as a disappointment. It may also come with the ring of truth; human cognition is what it is, and there are limits on what the law can do about it.<sup>150</sup> The persistence of the doctrines discussed in Parts II and III affords some evidence of how durable FAE can be.

What if, as I suspect, FAE is largely irreducible? On that assumption, FAE gives significant support to utilitarian, as opposed to retributive or virtue-based, theories of punishment.<sup>151</sup> In particular, FAE weighs heavily against mandatory retributivism, which would seem likely in practice to impose punishment that the theory itself forbids.

Retributive theory, which is probably the dominant model in American legal thought,<sup>152</sup> at least for the present,<sup>153</sup> regards

---

149. See generally ALAN DERSHOWITZ, *THE BEST DEFENSE* (1982).

150. Finkel, *supra* note 28, at 477.

The need to make attributions seems part and parcel of being human and endemic to science. And if the fundamental attribution error hangs on prototypical coattails, or is the tail that wags the dog, we best learn to understand it and how to more effectively rein it in, for eliminating it entirely seems out of the range of human capabilities.

*Id.*

151. For general surveys of the philosophy of punishment, see, for example, PHILIP BEAN, *PUNISHMENT: A PHILOSOPHICAL AND CRIMINOLOGICAL INQUIRY* (1981); *PHILOSOPHICAL PERSPECTIVES ON PUNISHMENT* (Gertrude Ezorsky ed., 1972) [hereinafter *PHILOSOPHICAL PERSPECTIVES*]; NIGEL WALKER, *PUNISHMENT, DANGER AND STIGMA* (1980).

152. See Dolinko, *supra* note 129, at 1623 ("It is widely acknowledged that retributivism, once treated as an irrational vestige of benighted times, has enjoyed in recent years so vigorous a revival that it can fairly be regarded today as the leading philosophical justification of the institution of criminal punishment." (footnote omitted)).

punishment inflicted for culpable wrongdoing as good of itself. Probably the most famous expression of this approach is Kant's dictum that even if a society were to disband the next day, its members should still execute a convicted murderer.<sup>154</sup> Modern retributivists give different accounts of precisely why punishment of culpable wrongdoing should be regarded as good of itself.<sup>155</sup> Retributivists disagree about whether blameworthy conduct affirmatively requires punishment or merely permits punishment if the balance of other considerations so inclines.<sup>156</sup> Despite these

153. Retributivism's stock has risen and fallen over time. See *id.* at 1623 nn.1-2; David Dolinko, *The Future of Punishment*, 46 UCLA L. REV. 1719, 1720 (1999) ("From our perspective today, we can see that those seemingly antiquated retributive notions of which Hart spoke have not only failed to disappear, but have come roaring back with—one might say—a vengeance.").

154. Immanuel Kant, *The Metaphysics of Morals* § 46(E), in KANT: POLITICAL WRITINGS 131, 156 (Hans Heiss ed., 2d ed. 1991).

Even if civil society were to dissolve itself with the consent of all its members (for example, if a people who inhabited an island decided to separate and to disperse to other parts of the world), the last murderer in prison would first have to be executed in order that each should receive his deserts and that the people should not bear the guilt of a capital crime through failing to insist on its punishment; for if they do not do so, they can be regarded as accomplices in the public violation of justice.

*Id.*

155. Kant and Hegel seem to have thought that respect for the offender's moral autonomy requires retributive punishment. See G.W.F. HEGEL, *THE PHILOSOPHY OF RIGHT*, excerpt reprinted as *Punishment as Right*, in PHILOSOPHICAL PERSPECTIVES, *supra* note 151, at 107 (retributive punishment "is regarded as containing the criminal's right and hence by being punished he is honoured as a rational being"); Kant, *supra* note 154, at 154-55 ("Judicial punishment can never be merely a means of furthering some extraneous good for the criminal himself or for civil society, but must always be imposed on the criminal simply because he has committed a crime. For a human being can never be manipulated just as a means of realising someone else's intentions, and is not to be confused with the objects of the law of kind."). Other theorists have justified retributive punishment on grounds of distributive justice, *i.e.*, the criminal has exploited the law-abiding by cheating on the law's system of reciprocal forbearance, so that fair play requires punishment, regardless of consequences. See JEFFRIE G. MURPHY, *RETRIBUTION RECONSIDERED* 15 (1992) ("Retributive theorists claim that punishment makes sure that wrongdoers suffer in proportion to their moral iniquity and thereby give up any unfair advantage over others their wrongdoing may have won them." (footnote omitted)); Herbert Morris, *Persons and Punishment*, reprinted in PHILOSOPHICAL PERSPECTIVES, *supra* note 151, at 116, 116-17 ("[I]t is just to punish those who have violated the rules and caused the unfair distribution of benefits and burdens. A person who violates the rules has something others have—the benefits of the system—but by renouncing what others have assumed, the burdens of self-restraint, he has acquired an unfair advantage. Matters are not even until this advantage is in some way erased.").

156. The view that culpable wrongdoing provides a necessary and sufficient reason for punishment is referred to as "positive retributivism," "pure retributivism," or "mandatory retributivism." The view that culpable wrongdoing authorizes, but does not require, punishment is referred to as "permissive retributivism." Both views have distinguished defenders. Kant appears to have taken the first view, although his discussion of the pardon power might be thought to qualify his views on punishment as a categorical imperative. See Kant, *supra* note 154, at 157 (discussing mitigation of death sentences when capital wrongdoing is so widespread

differences, all retributive theories share a family resemblance, rooted in the reciprocal ideas that punishment can be deserved, and thus it should never be undeserved.

Retributivism's main competitor is utilitarianism. The utilitarian approach, with its own impressive lineage running back to Beccaria<sup>157</sup> and Bentham,<sup>158</sup> regards punishment as an evil than can be justified only by a favorable balance of future consequences. If punishment prevents future crimes—by deterring others,<sup>159</sup> by restraining the offender,<sup>160</sup> by setting a moral example,<sup>161</sup> or by subjecting the offender to rehabilitation<sup>162</sup>—the suffering imposed on the guilty might have benefits exceeding the costs.

Utilitarians regard retributivism as little more than glorified revenge, and believe that characterizing the deliberate infliction of pain as a good in itself is perverse.<sup>163</sup> Retributivists criticize utilitarianism for reducing the offender to the status of a mere tool, a thing to be made an example of for the benefit of society.<sup>164</sup> They also note that utilitarianism logically permits punishment of the innocent, cruel punishments, strict liability, and anything else that might prevent future crime, however unfair or disproportionate such

---

that perfect justice might bring the state to “the stage of having no more subjects” or “blunt the peoples’ feelings by a spectacle of mass slaughter”). For more recent examples, see, for example, MICHAEL S. MOORE, *PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW* (1997) (defending mandatory retributivism); ANDREW VON HIRSCH, *DOING JUSTICE: THE CHOICE OF PUNISHMENTS* (1976) (defending permissive retributivism).

157. See generally CESARE BECCARIA, *ON CRIMES AND PUNISHMENTS* (1764) (Henry Paolucci trans., 1963) (influential early instrumental account of punishment, including defense of proportionality and criticism of capital punishment).

158. See generally JEREMY BENTHAM, *AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION* (1789) (seminal utilitarian account of punishment).

159. Deterrence plays a central role in economic analysis of criminal law. See, e.g., Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1193-95 (1985) (arguing that criminal law’s basic mission is to deter market bypassing behavior when tort sanctions may be ineffective).

160. See, e.g., JAMES Q. WILSON, *THINKING ABOUT CRIME* 198-201 (1975) (advocating a focus on incapacitation).

161. See, e.g., Dan M. Kahan, *Between Economics and Sociology: The New Path of Deterrence*, 95 MICH. L. REV. 2477, 2479 (1997) (defending an expressive account of punishment, including community policing, curfews, and alternative sanctions, designed to “deter as well as or better than severe prison terms and . . . cost much less”).

162. Rehabilitation, once trendy, has fallen out of style as a justification for punishment. See generally FRANCIS A. ALLEN, *THE DECLINE OF THE REHABILITATIVE IDEAL: PENAL POLICY AND SOCIAL PURPOSE* (1981).

163. See, e.g., H.L.A. HART, *PUNISHMENT AND RESPONSIBILITY* 234-35 (1968) (“To some critics it appears to be a mysterious piece of moral alchemy in which the combination of the two evils of moral wickedness and suffering are transmuted into good; to others the theory seems to be the abandonment of any serious attempt to provide a moral justification for punishment.”).

164. See, e.g., Kant, *supra* note 154.

measures might be.<sup>165</sup> Recent utilitarian rejoinders to these charges suggest that the debate is far from settled.<sup>166</sup> Perhaps motivated by a sense of the difficulties attending both the retributive and utilitarian positions, some theorists have taken an Aristotelian turn toward justifying punishment as a way of realizing virtue.<sup>167</sup> FAE poses an important challenge to all three types of theories, although I aim to show that the challenge is more serious for retributivists and virtue-ethics theorists.

## 1. Fundamental Attribution Error and the Judgments Required by Retributive Theory of Punishment

### *a. Types of Retributive Theories*

Any retributive theory of punishment must give an account of why deserved punishment is justified and how much punishment is deserved in particular cases.<sup>168</sup> Culpability-based retributivists believe that the actor's subjective awareness of wrongdoing triggers blameworthiness and makes the actor eligible for punishment.<sup>169</sup> Harm-based retributivists believe that causing or risking harm crosses the threshold of blameworthiness.<sup>170</sup>

Whatever the trigger of blameworthiness may be, punishing the blameworthy actor can be defended on various grounds. Some theorists see punishment as the logical corollary of respect for the offender's autonomy.<sup>171</sup> Others see punishment as fair play to the law-abiding, correcting the distributive injustice worked by the actor's

---

165. See, e.g., BEAN, *supra* note 151, at 44 (summarizing these weaknesses of utilitarian theories).

166. See generally Guyora Binder & Nicholas J. Smith, *Framed: Utilitarianism and Punishment of the Innocent*, 32 RUTGERS L.J. 115 (2000); Russell L. Christopher, *Deterring Retributivism: The Injustice of "Just" Punishment*, 96 NW. U. L. REV. 843 (2002).

167. See, e.g., Huigens, *supra* note 5.

168. H.L.A. Hart drew an influential distinction between the "general justifying aim" of punishment from the distribution of punishment. See HART, *supra* note 163, at 8-13. Any account of just punishment, whether consequentialist or deontological, must address both questions.

169. On the distinction between culpability-based and intention-based retributivism, see Cole, *supra* note 76, at 74-76.

170. This is probably the most common version of retributivism. Kant offers a leading illustration. Given Kant's ethical system, which focuses on the will of the actor rather than the consequences of the action, culpability is virtually built into the notion of retribution. Thus, Kant insisted that only an "undeserved evil" called for retribution. Kant, *supra* note 154, at 155. "Every transgression of the law can and must be explained only as the result of a maxim of the criminal whereby he makes a rule out of misdeeds like the one in question." *Id.* at 145 note \*.

171. See, e.g., HEGEL, *supra* note 155, at 107-08.

cheating on society's system of reciprocal restraint.<sup>172</sup> Yet another view is that retributive punishment is justified as a form of moral rhetoric, designed to persuade the criminal to renounce his wrongdoing and internalize the community's morality.<sup>173</sup> As to the measure of punishment, retributivists (at least outside the capital punishment context) appear to have abandoned the *jus talonis* in favor of some intuitive proportionality assessment.<sup>174</sup>

At least one additional important divide separates mandatory retributivism from permissive retributivism.<sup>175</sup> To mandatory retributivists, blameworthy behavior provides a necessary and sufficient reason for punishment. To permissive retributivists, blameworthy behavior only opens the moral door to punishment; its actual infliction might be withheld for sufficiently strong countervailing pragmatic reasons.

Within this schema there is widespread agreement that once the moral threshold of blameworthiness is crossed, two further conclusions are necessary before blame and punishment are justified. The observer must conclude that the actor was responsible for his behavior, in the sense that the result was not a faultless accident; *i.e.*, the observer must believe that the actor had some culpable state of mind. Finally, the observer must conclude that the actor has no justification or excuse for the behavior.<sup>176</sup>

---

172. See, e.g., Morris, *supra* note 155, at 116-18.

173. See R.A. Duff, *Penal Communications: Recent Work in the Philosophy of Punishment*, 20 CRIME & JUST. 1 (1996).

174. See, e.g., H.J. McCloskey, *A Non-Utilitarian Approach to Punishment*, reprinted in PHILOSOPHICAL PERSPECTIVES, *supra* note 151, at 119, 133-34.

175. See *supra* note 156.

176. See, e.g., KADISH, *supra* note 90, at 88; KELLY G. SHAVER, THE ATTRIBUTION OF BLAME 173 (1985) (arguing that the assignment of blame "presupposes a particular set of actions (those that produce negative consequences), a specific level of personal causality (single causation at the intentional level), a special combination of the dimensions of responsibility (causation, knowledge of the consequences, intentionality, voluntary choice, and the capacity to distinguish right from wrong), and the failure to have an adequate justification or excuse."); Thomas R. Schultz & John M. Darley, *An Information-Processing Model of Retributive Moral Judgments Based on "Legal Reasoning,"* in 2 HANDBOOK OF MORAL BEHAVIOR AND DEVELOPMENT 247, 256 (William M. Kurtines & Jacob L. Gewirtz eds., 1991) ("Blame is a decision that a person is at fault (given positive decisions on causation and responsibility)."). Kadish argues that

The three categories of legal excuses that I described in the previous section suggest the common rationale behind excuses in both the law and everyday moral judgments—namely, that justice requires the preclusion of blame where none is deserved. In the first category, the person is not to blame because he has no control over his movements; in the second, because though he breached a legal norm, he acted in circumstances so constraining that most people would have done the same; in the third, because though there is action in breach of a norm in circumstances where most people would *not* have done the

*b. Blameworthiness and FAE*

What all retributive theorists appear to share is a commitment to avoid punishment in the absence of blameworthiness. What does blameworthiness mean? The clearest approach to this question looks to the excuses; even when the actor's behavior was intentional and harmful, it may be excused on such grounds as insanity or duress.<sup>177</sup> Two basic approaches have emerged to define blameworthiness and explain why excusing situations reduce or nullify it.<sup>178</sup>

The first approach looks to the actor's conduct as evidence of his character.<sup>179</sup> If the actor caused harm in a situation in which a good person would not have caused harm, the harm is attributable to the actor's bad character. He may, on this character-based approach, properly be blamed for the character defect that caused the harm. An excuse, however, still can reduce or nullify the actor's blameworthiness when he has some extreme incapacity to understand or respect moral reasons for self-control.<sup>180</sup>

The second approach looks to the actor's opportunity to choose between causing harm and not causing it. These "choice theorists" consider the actor to be blameworthy even if his choice is an aberrant departure from his character. This seems plausible; the murder that breaks an otherwise consistent pattern of virtuous behavior is still blameworthy. If excuse is not to be reduced to justification, however, it is necessary for some voluntary choices that cause unjustified harm to be excused. The defendant acting under duress, for instance, acts

same, the person, because of a fundamental deficiency of mind, is not a responsible moral agent.

KADISH, *supra*, at 88.

177. Michael S. Moore, *Choice, Character and Excuse*, 7 SOC. PHIL. & POL'Y 28, 29 (1990) ("[T]he excuses are the royal road to theories of responsibility generally. The thought is that if we understand why excuse in certain situations but not others, we will have gained a much more general insight into the nature of responsibility itself.").

178. See, e.g., Rachael A. Hill, Comment, *Character, Choice, and "Aberrant Behavior": Aligning Criminal Sentencing with Concepts of Moral Blame*, 65 U. CHI. L. REV. 975, 1017-22, 1072-88 (1998) (summarizing choice and character positions and defending character approach).

179. See generally Peter Arenella, *Character, Choice and Moral Agency: The Relevance of Character to our Moral Culpability Judgments*, 7 SOC. PHIL. & POL'Y 59 (1990) (defending character approach).

180. Whether blame may properly attach to unchosen character traits is a deep and troubling question. See *id.* at 82-83. Even Professor Arenella, however, who would exclude from blame moral psychopaths who lack the ability to conform to the law for moral (as opposed to nonmoral) reasons, suggests that blame may be just so long as the actor has "some modest capacity for *critical self-reflection* and *self-revision*." *Id.* at 82. The extremity of the cases he finds troubling—the moral psychopath or the brainwashed accomplice—suggests that most criminal defendants have whatever capacity is required for making their departures from ordinary conduct blameworthy.

voluntarily and intentionally. The excuse is recognized not because there was an inability to exercise behavioral control, but because the choice presented was a hard one.<sup>181</sup> What makes some choices harder than others?

The most obvious answer is that the choice is hard, and an excuse should be recognized, when the situational pressure is so strong that many other persons would have responded by giving in to the pressure and causing the unjustified harm.<sup>182</sup> The accused behaved badly, but he was not deviant. Sometimes the law articulates the point of reference in normative terms (the reasonable person, the law-abiding person), and sometimes in statistically representative terms (the ordinary person or the person of ordinary firmness). If this focus on variance from a reasonable or ordinary person is what makes hard choices excusable, however, the choice theorist is really blaming the actor for his character.<sup>183</sup> Unlike the law's reasonable or ordinary person, the actor chose to give in to the situational pressure. He is blamed, in short, for having a different (and worse) character than other persons.

FAE has troubling implications for the retributivist's project of rationally assessing blameworthiness. The character-based approach directly embraces the project of inferring personality traits from behavior. This is the very inference that the psychological research suggests human observers will make too readily. Consider, in this regard, the Fidel Castro essays, the quiz master experiment, or the

---

181. See Moore, *supra* note 177, at 35 (defending excuse when the actor either lacked the capacity or the "fair opportunity" to act justifiably). The use of the word "fair" here is a tip-off that some normative idea of hard choices is being grafted on to the basic volitional account.

182. Professor Moore tries to avoid falling back on hypothetical reference persons to measure hard choices by arguing that hard choices are those where the choice of evils is a close call from an objective point of view. *Id.* at 40. There are at least two problems with this move. In the first place, the law actually articulates the excuses in terms of hypothetical reference persons. In the second place, to claim an excuse the actor must be motivated by the external pressure. An IRA gunman may credibly threaten the accused with murder for refusing to drive a getaway car; but if the accused is an IRA sympathizer who would be happy to help the gunmen even if unthreatened, the excuse is unavailable. We withhold blame from the excused actor because he is not deviant, not because he made a modest, as opposed to gross, arithmetic error.

183. See Hill, *supra* note 178, at 993.

Using a hypothetical reasonable person as a baseline introduces a concealed character judgment into the choice-based approach. As with Moore's "objectively regarded as evil" test, a reasonable person standard requires courts to assess whether an actor's reaction to a perceived threat or affront reveals a character trait that is undesirable according to society's standards. When an actor behaves unreasonably, he expresses a bad character and is, therefore, morally blameworthy. When an actor acts as a reasonable person would, he does not reveal a bad character and is excused.

*Id.*

foul shots taken in a dimly lit gymnasium.<sup>184</sup> In these experiments, observers held actors responsible despite the observers' knowledge of very serious situational constraints. Indeed the term "correspondence bias" refers precisely to the tendency to associate behavior with a corresponding trait.

In the choice approach, the problem recurs. FAE predisposes observers to exaggerate both volitional capacity and fair opportunity to resist situational pressure. A choice theorist who does not repudiate situational excuse altogether admits that some bad choices are not blameworthy. As a result of FAE, however, in deciding how hard a choice the actor faced, observers will tend to attribute the choice to the actor's character rather than the situation.

FAE tends to magnify the causal significance of the defendant's conduct relative to other factors. Observers predisposed to believe that the world is just need to identify personal, rather than impersonal, causes for negative events.<sup>185</sup> Compounding this tendency is the so-called hindsight bias, which inclines observers *ex post* to believe that actual events were probable *ex ante* even when they were not.<sup>186</sup> This, in turn, inclines observers to infer intention, knowledge, or recklessness from the foreseeability of events that were in fact not foreseeable.

Harm-based retributivists, with their focus on causing or risking harm, invite the tendency of observers to commingle fault with causation, amplified by the hindsight bias. A purely subjectivist culpability theorist, by contrast, considers the actor eligible for punishment based on his subjective awareness of wrongdoing. This may disadvantage the government unduly, as those who focus on the person rather than the situation interpret failed attempts as innocent accidents and harmless recklessness as due care.

As the utilitarians have pointed out, retributivists have some difficulty in determining the amount of punishment required by any given instance of culpable wrongdoing.<sup>187</sup> To the extent that retributivists rely on intuition or the sense of the community to measure proportionate punishments, FAE suggests that officials attempting to follow retributive theory will overpunish. Their

---

184. See *supra* notes 41-62 and accompanying text.

185. On the just-world hypothesis, see generally M. LERNER, *THE BELIEF IN A JUST WORLD* (1980). On the implications of just-world beliefs for attributions of blame, see FEIGENSON, *supra* note 3, at 100-101 (explaining how just-world thinking can lead to blaming the victim).

186. See, e.g., FEIGENSON, *supra* note 3, at 62-64.

187. See, e.g., BEAN, *supra* note 151, at 29 ("Retribution is unable to provide clear guidelines as to what equivalence should mean in practice."). See generally Dolinko, *supra* note 129.

intuitions will tend to overassess personal as opposed to situational factors at the time of the wrongdoing.

This is especially likely for those retributivists who root the justification of punishment in distributive justice. On that account, the criminal deserves punishment for deviating from society's scheme of mutual restraint. Yet if many people would have acted just as the defendant did given the same situation, the defendant hasn't really helped himself to any unfair advantage. If that really is why punishment is deserved, retributivism in practice will dish out far more punishment than fair play can justify.

All of these tendencies are of much greater concern for mandatory retributivists than for their permissive counterparts. An exaggerated sense of blame that only authorizes, but does not require, punishment is little different in practice than a utilitarian theory subject to (likewise exaggerated) culpability-based side-constraints. But for the mandatory retributivist in the Kantian tradition, FAE is a major challenge. The mandatory retributivist can hold his ground only in the belief that legal procedures can counteract cognitive predispositions. Yet, as we have seen, in many ways the legal system has failed to counteract FAE. Nor is it entirely clear just what procedures might do so. Given the high priority retributivists place on not punishing the innocent, the failure of the legal system to counteract cognitive predispositions is a strong point against operating an actual criminal justice system according to mandatory retributivism's prescriptions.

The debate over theories of punishment often seems little more than a clash of intuitions. For example, in a prominent recent defense of using consequentialist rather than deontological norms in policy making, Professors Kaplow and Shavell point out that deontological theories have the inherent capacity to justify policy choices that make everyone worse off.<sup>188</sup> They seem to think that this turn to the Pareto principle makes a decisive rhetorical point—deontologists will agree that policies that make everyone worse off are indefensible.

This rhetorical move, however, has less persuasive power than its authors suppose. Consider, for instance, a research physician convicted of murder and sentenced to death under a given jurisdiction's prevailing law. Other experts credibly aver that the convict has unique gifts and is on the threshold of major advances in combating cancer or some other dread disease. Commuting the sentence makes the doctor better off. It would also make all at risk of

---

188. Louis Kaplow & Steven Shavell, *Fairness Versus Welfare*, 114 HARV. L. REV. 961, 1011-17 (2001).

the disease (and all who hold those at risk in some affection) better off. Without expressing an opinion on the merits, I am fairly confident that many reasonable minds would swallow the Paretian pill and say "*fiat justitia, coelum ruit.*"<sup>189</sup>

By contrast, the psychological research invokes facts, not values. Retributive theory basically asks human beings to characterize the behavior of other human beings as harmful, responsible, and inexcusable. Psychological research indicates that these judgments will reflect the predispositions of those doing the judging as well as genuine phenomena. I can imagine a determined retributivist saying that *some* conduct deserves blame and that *some* institutional arrangements can be devised to accurately identify these cases. But without great confidence in these institutional arrangements, retributive theory as applied by human judgment is very likely to produce punishment in excess of its theoretical justification.

## 2. Virtue-Based Theories of Punishment

FAE also has troubling implications for theories of punishment based on the promotion of virtuous character.<sup>190</sup> A character-based virtue-ethics theory, leaving FAE aside for the moment, runs into the debate, now ongoing among philosophers, about whether the person/situation research forces us to surrender the very concept of character traits, at least in the broad terms they have been

---

189. Cf. Kant, *supra* note 154, at 155.

The penal law is a categorical imperative, and woe betide anyone who winds his way through the labyrinth of the theory of happiness in search of some possible advantage to be gained by releasing the criminal from his punishment or from any part of it, or who acts in the spirit of the pharisaical saying: "It is better that one man should die than that the whole people should go to ruin." For if justice perishes, there is no further point in men living on earth. What then are we to think of the proposal that the life of a condemned criminal should be spared if he agrees to let dangerous experiments be carried out on him in order that the doctors may gain new information of value to the commonwealth, and is fortunate enough to survive? A court of justice would dismiss with contempt any medical institution which made such a proposal; for justice ceases to be justice if it can be bought at a price.

*Id.*

190. Aristotle believed that a system of official punishment was necessary to foster virtuous character. See ARISTOTLE, *ETHICS* 336-37 (J.A.K. Thompson trans., rev. ed., Penguin Books 1976). Responding to the objection that punishing character deficiencies is unjust because character traits are not voluntarily chosen, Aristotle argued that at least some character flaws are self-inflicted and thus justly subject to punishment. See *id.* at 124-26. For a modern defense of this view, see Edmund L. Pincoffs, *Legal Responsibility and Moral Character*, 19 WAYNE L. REV. 905 (1973).

understood since Aristotle.<sup>191</sup> The psychologists have cast serious doubt on the very existence of any cross-situational disposition as broad as Blackstone's "vicious will."<sup>192</sup> A virtue theory that did describe traits defined narrowly enough to be recognized in the psychological research, however, would have to account for an unmanageably large number of predispositions. Imagine the virtue theorist saying, "You are being punished for your unvirtuous propensity to steal coins out of parking meters, at night, when unwatched by the police, drunk, and broke."

The thief's bad character may not run any deeper or wider than that predisposition. To treat such precisely defined dispositions as proper predicates for punishment, however, would essentially reduce the character-based scheme to a naked system of social control. Moreover, a great many law-abiding people are simply lucky enough to have avoided similarly precise situational triggers of bad behavior. The focus on the offender's character therefore seems a highly inappropriate basis for coercive intervention by the state.

Whether virtue-ethics of the character sort can support a theory of punishment given the person/situation research, however, is beside the point. A virtue-ethics approach to punishment depends not only on the existence of character traits, but also on the ability of observers to identify and evaluate traits. Put into practice, the result would be systematically biased in favor of blaming people for character traits they don't have. Observers committing FAE will classify one who steals from parking meters at night while drunk and broke simply as a thief or still more generally as dishonest. But the disposition to steal from unguarded parking meters at night while drunk is not the same as a predisposition to cheat on one's taxes or to burglarize residences.

Any move to replace the inquiry into bad character with an inquiry into the practical reasonableness of the accused's behavior, all things considered,<sup>193</sup> might well be still worse. Such an approach relies

---

191. Consider how broadly Aristotle defined the virtues and vices—"courage," as distinct from "rashness" on the one hand or "cowardice" on the other, and so on. See ARISTOTLE, *supra* note 190, at 104 (table of virtues and vices). Or consider how Professor Pincoffs describes the relevant character traits: "We have a moral right to demand—each of us of every other—honesty, trustworthiness, justice, and other qualities as well." Pincoffs, *supra* note 190, at 919.

192. See *supra* text accompanying notes 34-40.

193. If I understand him correctly, Professor Huigens favors this approach. See Kyrøn Huigens, *Virtue and Inculpation*, 108 HARV. L. REV. 1423, 1437-38 (1995).

The short answer is that inculpation, as an inquiry into a person's relation to and responsibility for the greater good, is an inquiry into the person's character—more particularly into the soundness, maturity and breadth of her practical judgment, the faculty by which she assembles her conception of the

very heavily on the tribunal's intuitive sense of justice.<sup>194</sup> Social cognition research teaches that such intuitive assessments of blameworthiness are likely to be biased in favor of identifying the choices of the accused, rather than the context, as the key link in the causal sequence.<sup>195</sup>

In a fascinating illustration of FAE in operation, Professor Huigens, defending a virtue-ethics approach, argues for full consideration of battered-women's syndrome evidence.<sup>196</sup> Here, with another human being to blame,<sup>197</sup> Professor Huigens can see some room for excuse; at the same time he categorizes criminal behavior by the addicted<sup>198</sup> or the unlucky or the stupid<sup>199</sup> as culpable failures of practical reasonableness. He is, as we have seen, far from alone in

good and her scheme of ends. If we take virtue to be, at bottom, this sort of exemplary practical judgment, then inculcation is a judgment on virtue.

*Id.*

194. See *id.* at 1462-65 (discussing "Virtue as a Jury Question").

195. Professor Huigens acknowledges the legal constraints on jury decision making, but nonetheless maintains that "[i]n the hard case, the jury acts as I have described it above: each member comparing the accused's choices with what she believes her own would be in the situation of the accused." *Id.* at 1466. If the psychologists deliberately set out to frame an inquiry which would give FAE maximum influence, they could hardly do better.

196. Kyron Huigens, *Street Crime, Corporate Crime, and Theories of Punishment: A Response to Brown*, 37 WAKE FOREST L. REV. 1, 23-28 (2002).

197. Professor Huigens begins a major article with this quotation, attributed to a prison warden in Michigan:

Frequently, someone who's committed pre-meditated murder, which in Michigan is how you get life with no possibility of parole, is not a good problem solver. They have elected to kill someone as a solution to a problem that they've identified. But now that that person is dead, the problem is resolved. . . .

"I caught a case." It's like they went fishing and it just leaped on their line or something. They didn't commit a crime. They don't frame it that way. They're a victim. And it's incredible to me when you know the kinds of crimes that some of these fellows have committed, that they still do not see the person that they brutalized as the victim. They see themselves as a victim.

Huigens, *supra* note 193, at 1423 (citation omitted). Note how both the warden and Professor Huigens easily pick up the attribution error of the murderers. Both the professor and the warden then turn about and attribute crime to person rather than situation. The murderer is a poor "problem solver" who "elected to kill." The construction of blame-worthy character defects from the fact of antisocial behavior is easy, perhaps even natural. The psychological research, however, suggests that this tendency to blame people rather than circumstances is at best exaggerated.

198. See Huigens, *supra* note 196, at 20-21 ("Because it breaks down the hard distinction between retribution and the pursuit of social welfare, the aretaic theory of punishment enables us to see drug treatment courts as an integrated part of the punishment system, and not as an alternative to punishment that necessarily slights desert and retribution.").

199. See Huigens, *supra* note 193, at 1476 ("Inculcation is premised on that responsibility for the good, not on the more narrow responsibility to avoid harm. One can therefore act culpably while being unaware of the particular risk one has created.").

embracing the “abuse excuse” while giving short shrift to situational excusing circumstances such as necessity and intoxication.<sup>200</sup>

### 3. Utilitarian Accounts

FAE complicates matters for utilitarians as well. The standard counter to the charge that utilitarianism permits punishing the innocent, cruel punishments, or strict liability, is that any system that permitted such practices would do more harm than good and therefore be inconsistent with utilitarianism.<sup>201</sup> Thus the utilitarian theorist typically builds such ideas as culpability and proportionality into the theory as side-constraints limiting the pursuit of overall utility in an individual case.<sup>202</sup> If culpability and proportionality are likely to be applied badly in a retributive system, these concepts are just as vulnerable to FAE when incorporated into utilitarian accounts as side-constraints.

The utilitarian theory of punishment may still have an advantage in practice. Retributive theory, at least in its mandatory form, regards punishment as a positive good; utilitarian theory requires that punishment be justified by expected future returns. To the extent that observers systematically tend to exaggerate the degree of culpable wrongdoing, retributivism inherently tends toward punishment that is excessive by whatever rational criteria any particular retributivist might select. Considerations of utility, however, might frequently point in the direction of punishing less than retributive theory might require. For example, a utilitarian skeptical about the marginal deterrent effect of executions and concerned about the cost and delay attending the capital-sentencing process might very well oppose the death penalty while a retributivist in the mold of Kant might support it.<sup>203</sup>

A utilitarian theory that incorporates retributive principles as side-constraints is vulnerable to punishing as much as retributive theory permits, but no more. The culpability side-constraint will be expanded by FAE, but within the constraint, the utilitarian will be seeking ways to inflict less pain on offenders so long as deterrence and restraint are adequately served. By contrast, retributive theory, applied by observers peering through the magnifying glass of FAE,

---

200. See *supra* notes 117-122 and accompanying text.

201. See, e.g., HART, *supra* note 163, at 8-13; John Rawls, *Two Concepts of Rules*, 64 PHIL. REV. 3 (1955).

202. See HART, *supra* note 163, at 8-13.

203. Compare Kant, *supra* note 154, with BECCARIA, *supra* note 157.

may call for “proportionate” punishment for culpable wrongdoing even when no social advantage is thereby gained.

To take but one example, there is general agreement that the sexual abuse of children is a terrible thing that society has every right to prevent, by force if necessary. In these cases, however, the divide between the two theories can be stark. Given the seriousness of the offense, retributivism counsels terrible penalties. The utilitarian, however, is free to ask whether long-term restraint *outside* the regular prison system might satisfy the needs of social control at less cost, in both dollars and misery.

More generally, utilitarian theories seem to fit the facts of human behavior better than retributive theories. The utilitarian intends punishment to change future situations by deterring or restraining harmful conduct. The utilitarian’s primary focus is on the future, not the past. She will administer punishment with an eye on how persons other than the defendant will respond to an environment in which they may expect to be punished for similar conduct.

On some debates, one does not really expect to change many minds. Still, this latter attitude—of inflicting the least amount of punishment that will serve the purposes of social control, in a spirit of sad necessity—comports better with what we know of situational determinants of behavior and the tendency to ascribe behavior to personality. If modern psychological research correctly describes these phenomena, the retributivists’ deliberate infliction of pain in an open spirit of moral superiority will be counterfactual in a great many cases.

Of course, if it were possible to neutralize cognitive biases in assessing blame, the theoretical debate would remain in its present state. But if both families of justifications for punishment run a serious risk of overassessing individual blameworthiness, it might be time to move the theoretical debate away from what a perfect theory of punishment requires, and toward the design of institutions that would minimize the irrational application of whatever theory of punishment prevails at the moment.

## VI. ANTICIPATING TWO PLAUSIBLE CRITICISMS

In other contexts, two plausible criticisms have been made of the use of psychological research for purposes of analyzing legal rules and institutions. One criticism has been to interpret the turn to behavioral science as an attempt to qualify the microeconomist’s rational-actor assumption, and to defend traditional law-and-

economics analysis against this perceived attack.<sup>204</sup> The other has been to point out that both behavior and cognition have multiple and complex causes, that it is a mistake to abstract a particular cognitive bias or behavioral tendency from others, and that lawyers are too quick to pounce on research findings that can be adapted to ulterior doctrinal purposes.<sup>205</sup> I shall refer to the first criticism as the “rational-actor assumption” and the second as the “law-office psychology criticism.”<sup>206</sup>

Whatever the merits of the rational-actor assumption in other contexts (and they are considerable),<sup>207</sup> this Article has dealt with decisions made by officials in the criminal-justice system. The system makes a major and substantially successful effort to distance these officials from considerations of immediate self-interest that rational-actor assumptions generally consider to motivate individuals. Bribes are forbidden, federal judges serve for good behavior, state judges serve long terms, and jurors in the vast bulk of cases have no stake in the outcome worth considering. This is not to say that rational-actor assumptions don’t have explanatory power in the criminal-justice context, only that the power of self-interest to offset other impulses will be less in this context than in others.

Moreover, there are many strands of current criminal-justice doctrine that are hard to explain solely on rational-actor grounds. Legislators, presumably following or sharing the views of their constituents, generally declined the Model Penal Code’s opportunity to equate attempt with success—an equation tailor-made for law-and-

---

204. See generally Richard A. Posner, *Rational Choice, Behavioral Economics, and the Law*, 50 STAN. L. REV. 1551 (1998) (defending rational-actor assumption against behavioral economics); Robert E. Scott, *The Limits of Behavioral Theories of Law and Social Norms*, 86 VA. L. REV. 1603 (2000) (arguing that modifying rational-actor assumption may deepen analysis but reduce its predictive power).

205. See Mitchell, *supra* note 21, at 1912 (“When the full range of empirical research into judgment and decision making is considered, and when the methodological assumptions and choices of this research are laid bare, it becomes clear that this body of research does not present an unqualified account of pervasive and systematic irrationality.”).

206. The phrase is a play on the well-known jibes of professional historians against “law office history.” See, e.g., Martin H. Flaherty, *History “Lite” in Modern American Constitutionalism*, 95 COLUM. L. REV. 523, 525 (1995) (“[C]onstitutional discourse is replete with historical assertions that are at best deeply problematic and at worst, howlers.”); Alfred H. Kelly, *Clio and the Court: An Illicit Love Affair*, 1965 SUP. CT. REV. 119, 122-32.

207. See, e.g., Donald A. Dripps, *Criminal Procedure, Footnote Four, the Theory of Public Choice; Or, Why Don’t Legislatures Give a Damn About the Rights of the Accused?*, 44 SYRACUSE L. REV. 1079 (1993) (arguing that public choice theory, founded on rational-actor assumption, accurately predicts legislative hostility to rights of the accused); Donald A. Dripps, *Living With Leon*, 95 YALE L.J. 906 (1986) (arguing that costs of obtaining search warrant, independent of deterrent influence of exclusionary rule, give rational police strong disincentive to seeking warrants not supported by probable cause).

order politicians and a godsend to prosecutors seeking extra leverage in plea bargaining.<sup>208</sup> Nor does self-interest, springing from fear of crime, explain the narrowing of the insanity defense following the Hinckley case,<sup>209</sup> given that civil commitment follows acquittals. If politicians were responding to widespread political passion in that instance, what explains the political passion? Not, I submit, the risk that waves of gunmen would now dare to murder in the hope that, once acquitted, each could spend the rest of his days at St. Elizabeth's. The examples could be multiplied, but I think these two give some evidence that FAE may have something to do with the shape of the law.

As to the law-office psychology objection, I am in no position to deny it. Either I have digested the psychological literature more-or-less correctly, or I haven't, and only those better versed in that literature can assess the degree to which I have or haven't. In the range of cognitive strategies recognized in the literature, it is hard to think of one that would oppose the operation of FAE with any consistency, and if there were such a strategy it should have shown up in the classic experiments on attribution. Moreover, FAE, unlike some of the other biases studied in the literature, seems to have a substantial effect size.<sup>210</sup> Still, I would be a poor scholar indeed if I didn't advise the reader that I am an academic lawyer, not a psychologist (or an economist, or a moral philosopher, or a historian—though I have found taking a position on economic, philosophical, and historical issues unavoidable in prior work). Nor has my argument aimed at any particular policy conclusion; it may be wrong, but it is ingenuous.

One more point deserves to be made about the law-office psychology objection. To the extent that FAE does have explanatory power about criminal justice, it suggests that my interpretation of the psychological dynamics is not misplaced. If FAE helps to explain legal practices that otherwise are hard to explain, those practices in turn tend to confirm that observers are inclined to attribute conduct and its consequences to the person rather than the situation.

## VII. CONCLUSION

This Article has not endorsed (or opposed) any specific doctrinal reforms. Instead, it has sought to explore how pervasive

---

208. See *supra* notes 69-71 and accompanying text.

209. See *supra* notes 103-111 and accompanying text.

210. See *supra* notes 49-57 and accompanying text.

tendencies of human cognition might affect official decisions in the criminal justice process. Multiple influences bear on each of these choices, and many criminal justice decisions, such as those made by legislators, are collective rather than individual. It follows that FAE is unlikely to be a necessary and sufficient explanation for any particular rule or practice in the system.

FAE does have at least one, highly general, doctrinal implication. We should be wary of defining criminal liability in terms of standards that appeal to intuitive judgments about personal responsibility. Human beings are predisposed to give affirmative answers to questions about personal responsibility. This is not to say that standards might never be the best formulation, all things considered; but it does suggest that lawmakers should leave catastrophic judgments of blameworthiness to intuition only when more specific formulations are clearly unworkable.

FAE helps to explain a great many puzzles in criminal law, such as the success/attempt distinction and the relative success enjoyed by the "abuse excuse." FAE also sheds some light on the theoretical divide between retributive and utilitarian accounts. On that front, I have expressed a certain sympathy for the utilitarian approach, which seems far more consistent with the evidence about human behavior and less susceptible to the risk of translating cognitive tendencies into years in the penitentiary.

The theoretical debate, moreover, may very well be seen as of much greater practical importance in light of FAE. Many actors in the criminal justice process come to a pragmatic accommodation of utilitarian and retributive theories after repeatedly finding that they rarely conflict in actual cases. What, however, if this happy coincidence turns out to be the product of irrational overassessments of blameworthiness? In that case, considerations of social control would counsel harsher punishment than rational considerations of retributive justice allow. FAE might be functional, precisely because it permits useful behavior modifications that ingrained notions of justice would forbid if rationally applied. If that is so, and if it were possible to counteract FAE across the system, we would need to resolve the theoretical dispute before we could decide important policy issues. Given the stubbornness of FAE it may well be a long time before the peaceful coexistence of utilitarian and retributive theories in practice is disrupted.

It may turn out, then, that not all of our predispositions are dysfunctional. Consider how FAE magnifies the satisfaction attending personal achievement, thereby maximizing pro-social incentive effects. Upon concluding prior articles I have taken a certain pride (no more

than most academics, I expect, but a healthy measure nonetheless) in having advanced a novel and plausible approach to some challenging problem. This Article is different. In this case, I must attribute my contribution to the coincidence that, unlike most teachers of criminal law, I also teach evidence, a field in which FAE has been identified as a possible justification for the character evidence rules.<sup>211</sup> In light of this somewhat peculiar situation, it seems unsurprising (and unimpressive) that I should have connected the psychological research current in the evidence literature to recurring issues in criminal law.

Thinking along these lines really might undermine an individual's—or a system's—commitment to the free-will assumption. Success and failure alike might be seen as undeserved, a characterization that would discourage personal forbearance in the pursuit of valued ends. To have this effect, however, the self-conscious recognition of FAE would need to go beyond what the evidence supports. Empirical evidence suggests that we overestimate the causal significance of personal forbearance, but it does not teach that self-control makes no difference at all. Blame and praise are more than useful conventions.

At any rate, as between the risks that accompany pretending to exercise a fuller rationality than we possess, and the risks that accompany admitting the probable distortions in our judgments of blame and praise, the balance inclines in favor of persevering in the direction of fully rational judgments.<sup>212</sup> In the actual practice of criminal justice, we frequently give legal force to intuitive judgments of personal responsibility, and this practice is approved by many strands of both popular theory and prevailing practice. But our intuitions about blame are not to be trusted. What criminal law reformers can learn from the research on FAE—and it may be all we can learn from it—is that we should reach judgments of blame with as much humility as possible.

---

211. This assumes that the piece has at least some positive value. Should it become an embarrassment—held up to the untenured as a model of the risks of writing out-of-field or in an interdisciplinary way without the right background—I shall probably come to the view that my colleagues should have convinced me not to publish it, that the VANDERBILT LAW REVIEW editors should have rejected it, or whatnot. Cf. *supra* notes 58-59 and accompanying text.

212. Cf. Posner, *supra* note 204, at 1575 (“One might have thought that behavioral economics had at least one clear normative implication: that efforts should be made through education and perhaps psychiatry to cure the cognitive quirks and weakness of will that prevent people from acting rationally with no offsetting gains.”).