Journal of Criminal Law and Criminology

Volume 113
Issue 1 *Winter*

Article 2

Winter 2023

Habit, Crime, and Culpability

Eric A. Johnson
University of Illinois College of Law

0091-4169/23/11301-0035
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY Vol. 113, No. 1

Printed in U.S.A.

HABIT, CRIME, AND CULPABILITY

ERIC A. JOHNSON*

Courts and scholars long have distinguished the wrongdoing component of criminal liability from the culpability component. In the old days, wrongdoing was thought to be crime's physical, objective component—the "evil-doing hand." Culpability, by contrast, was the mental, subjective component—the "evil-meaning mind." Nowadays, most scholars agree with Holmes that even the wrongdoing component requires proof of the actor's mental state. If the wrongdoing component requires proof of the actor's mental state, though, what's the point of the culpability requirement? For now, the dominant answer appears to be that the culpability requirement is a concession to human weakness.

In this Article, I will develop a different view. I will argue that the culpability requirement is less a concession to human weakness than to the varieties of human rationality. Building on insights by philosopher Michael Bratman and others, I will argue that rationality can take at least two fundamentally different forms. The wrongdoing requirement is concerned only with conduct's time-slice rationality—with the act's downstream risks and utilities as measured from the moment of the act. Conduct that isn't time-slice rational, however, still can embody a second kind of rationality, namely, temporally extended rationality. This second variety of rationality is present, for example, when an actor's conduct is attributable to desirable habits of thinking, feeling, or behaving. The culpability requirement is best understood as addressed to this second kind of rationality. It absolves just those actors whose conduct, though wrongful, nevertheless is a product of desirable habits.

* Professor of Law and Heidi Hurd Faculty Scholar, University of Illinois College of Law.
I am grateful to Ernest Johnson for his comments on an earlier version of this paper.

INTRODUCTION

Culpability is a mystery. If you ask a lawyer what culpability is, they'll probably say something about so-called culpable mental states.[1] But this isn't right. Mental state requirements in criminal statutes are mostly designed to ensure that the actor's conduct is objectively wrongful, not that it is culpable.[2] Wrongfulness, after all, usually is about the risk posed by an actor's conduct.[3] And the risk posed by an actor's conduct is measured according to "the circumstances known to him," in Holmes's and the Model Penal Code's definitive formulation.[4] To decide whether the actor's conduct is wrongful,

---

[1] *See, e.g.*, Celis v. State, 416 S.W.3d 419, 435 (Tex. Crim. App. 2013) (Keller, J., concurring) (explaining that the part of the Texas Penal Code entitled "Requirement of Culpability" is "devoted exclusively to the treatment and application of culpable mental states").

[2] *See* Eric A. Johnson, *Rethinking the Presumption of Mens Rea*, 47 WAKE FOREST L. REV. 769, 786–88 (2012) (explaining why the so-called mens rea presumption really is "an actus reus presumption").

[3] OLIVER WENDEL HOLMES, JR., THE COMMON LAW 75 76 (1881).

[4] *Id.*; MODEL PENAL CODE § 2.02(2)(c), (d) (AM. L. INST. 1985); *see generally* Eric A. Johnson, *Knowledge, Risk, and Wrongdoing: The Model Penal Code's Forgotten Answer to the Riddle of Objective Probability*, 59 BUFF. L. REV. 507 (2011) (explaining and defending Holmes's and the Model Penal Code's "known circumstances" formula for calculating objective probabilities).

then, the factfinder has to determine what the actor knew.[5] It is mostly to this end that statutes require proof of the actor's mental states.[6]

The fact that the conduct's wrongfulness is judged from the actor's perspective is part of what makes the separate culpability requirement so mysterious.[7] If the wrongdoing component of criminal liability already requires proof that the actor knew the very facts in which the conduct's risk inhered, why would the law require anything more? What function is left for the culpability component to perform?[8]

The usual answer to this question—that the culpability requirement operates as "a concession to human weakness"—is frustratingly vague.[9] This vagueness is reflected, for example, in what the Model Penal Code says about culpability. The Code's commentary acknowledges that the recklessness and criminal negligence standards require the factfinder to make a separate "culpability judgment," over and above its wrongfulness judgment.[10] But

---

[5] In crimes of recklessness and criminal negligence, the law requires the factfinder to calculate the risk on the basis of the "circumstances known to [the actor]." MODEL PENAL CODE § 2.02(2)(c), (d). Accordingly, as some courts have recognized, the government must prove the actor's "awareness of the attendant circumstances leading to such a risk," whatever those circumstances might be. Queeman v. State, 520 S.W.3d 616, 623 (Tex. Crim. App. 2017) (quoting Montgomery v. State, 369 S.W.3d 188, 193 (Tex. Crim. App. 2012)). By contrast, other criminal statutes require the government to prove the actor's awareness of specific *kinds* of facts. Drunk-driving statutes, for example, may require the government to prove that the actor knew he or she was driving and knew he or she had been drinking. *See* State v. Simpson, 53 P.3d 165, 167 (Alaska Ct. App. 2002) (comparing Montana and Alaska drunk-driving statutes). In either case, though, the purpose of the mental state requirement is the same: to prove the actor's knowledge of the facts in which the risk inhered.

[6] Johnson, *supra* note 2, at 787.

[7] *See* Kent Greenawalt, *The Perplexing Borders of Justification and Excuse*, 84 COLUM. L. REV. 1897, 1907–09, nn.30–32 (1984) (exploring the roles of factual mistakes in justification and excuse).

[8] *Cf.* Kenneth W. Simons, *Dimensions of Negligence in Criminal and Tort Law*, 3 THEORETICAL INQUIRIES L. 283, 299 (2002) (acknowledging that "the *ex ante* perspective that is normally a necessary feature of the tort negligence judgment itself presupposes a certain kind of 'cognitive' [negligence]").

[9] *E.g.*, Vera Bergelson, *Justification or Excuse? Exploring the Meaning of Provocation*, 42 TEX. TECH L. REV. 307, 317 (2009) ("The defense of duress is a limited concession to human weakness, namely, fear of death or serious bodily harm."); Richard A. Bierschbach & Alex Stein, *Mediating Rules in Criminal Law*, 93 VA. L. REV. 1197, 1247–48 (2007) ("To the retributivist, excuses are concessions to human frailty that a liberal society gives to individuals whose conduct it still deplores."); Alan Brudner, *Excusing Necessity and Terror: What Criminal Law Can Teach Constitutional Law*, 3 CRIM. L. & PHIL. 147, 157 (2009) (summarizing the "concession to human frailty" theory in relation to "excusing necessity").

[10] MODEL PENAL CODE & COMMENTS. § 2.02 cmt. 3 at 238 (AM. L. INST., Official Draft and Revised Comments 1985) [hereinafter MPC COMMENTARIES].

when it comes time to say what this culpability judgment consists of—to say exactly what distinguishes culpable wrongdoing from wrongdoing that merely is expressive of ordinary human fallibility—the commentary punts: "There is no way to state this value judgment that does not beg the question in the last analysis; the point is that the jury must evaluate the actor's conduct and determine whether it should be condemned."[11] The Code takes the same approach to culpability-negating "excuse" defenses like duress and extreme emotional disturbance.[12] The commentary says, in effect, that both defenses require factfinders merely to decide whether they are able to identify with the actor despite his or her wrongdoing.

In this Article, I will challenge the view that the culpability requirement is a concession to human weakness. Building on insights by philosopher Michael Bratman and others, I will argue that the culpability requirement is, rather, a concession to the varieties of human rationality. The wrongdoing component is addressed to one sort of rationality, which I will call time-slice rationality.[13] The culpability component is addressed to another, namely temporally extended rationality.

Time-slice rationality, as embodied in the criminal law's wrongdoing component, is about the downstream risks and benefits of the actor's conduct under "the circumstances known to him."[14] To decide whether the actor's conduct is wrongful, the factfinder applies a criminal law variant of the so-called Hand formula.[15] And in applying this Hand formula variant, the

---

[11] *Id.* at 237.

[12] *Id.* § 2.09 cmt. 2 at 374–75 (explaining that the duress defense is designed to absolve defendants in cases where the factfinder is "not prepared to affirm that they [would behave differently than the defendant] if their turn to face the problem should arise."); MODEL PENAL CODE & COMMENTS. § 210.3 cmt. 5(a) at 63 (AM. L. INST., Official Draft and Revised Comments 1980) ("In the end, the question [posed by the partial defense of extreme mental or emotional disturbance] is whether the actor's loss of self-control can be understood in terms that arouse sympathy in the ordinary citizen.").

[13] *See* MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 78 (1999) (distinguishing the "time-slice standard" for assessing rationality from "the model of temporally extended rational agency").

[14] MODEL PENAL CODE § 2.02(2)(c), (d) (AM. L. INST. 1985).

[15] Joshua Dressler, *Does One Mens Rea Fit All?: Thoughts on Alexander's Unified Conception of Criminal Culpability*, 88 CALIF. L. REV. 955, 957 (2000) ("To determine justifiability [in connection with recklessness], we conduct a criminal law version of the Learned Hand formula for measuring civil negligence . . . ."); George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537, 542 n.19 (1972) (observing that Model Penal Code § 3.02 uses a variant of the Hand formula to define the choice-of-evils defense); Leo Katz, *A Look at Tort Law with Criminal Law Blinders*, 76 B.U. L. REV. 307, 308 (1996) ("The drafters of the Model Penal Code seem to have been so taken by this claim [that the Hand formula clearly captures our intuitions about the meaning of negligence] as to adopt a

factfinder treats the actor as a "time-slice agent," who calculates the act's risks and benefits "from scratch" in the instant before the act.[16] In effect, the factfinder treats the actor as if the actor, in that instant, had thought about all the possible causal sequences that might be set in motion by the conduct, had assigned probabilities to these causal sequences on the basis of "the circumstances known to him,"[17] had assigned values to possible outcomes, and finally had aggregated these values and balanced the cumulative risks against the benefits.

Everybody realizes, of course, that actors rarely if ever engage in the sort of full-blown deliberation that is presupposed by the Hand formula.[18] In reality, some or all of this idealized deliberative process usually will be preempted by the actor's habitual ways of thinking, feeling, or behaving.[19] Sometimes, behavioral habits will short-circuit the deliberative process entirely; they will cause the actor to perform a bodily movement "in a reflexive, automatic manner" without any conscious deliberation at all.[20] Even when the actor *does* consciously deliberate before acting, habits of thought or feeling will short-circuit aspects of the deliberative process.[21]

---

formulation pretty close to it: They define negligence as the taking of a substantial, unjustifiable risk.").

[16] BRATMAN, *supra* note 13, at 79.

[17] MODEL PENAL CODE § 2.02(2)(c), (d) (AM. L. INST. 1985).

[18] *See* JENNIFER K. ROBBENNOLT & VALERIE P. HANS, THE PSYCHOLOGY OF TORT LAW 4 (2016) ("The psychological research . . . makes clear that risk-utility balancing in the strictest sense does not completely capture the nature of decision making about risk."); CHARLES DUHIGG, THE POWER OF HABIT xv–xvi (2012) ("Most of the choices we make each day may feel like the products of well-considered decision making, but they're not. They're habits."); *cf.* Brad Hooker, *Rule Consequentialism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY, https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule [https://perma.cc/9JT2-FG52] (observing that consequentialists almost never defend a purely "act-consequentialist decision procedure as a general and typical way of making moral decisions").

[19] ROBBENNOLT & HANS, *supra* note 18, at 41–42 ("Decision making about risk frequently happens quickly and unconsciously and is affected by a variety of heuristics.").

[20] Wendy Wood, Asaf Mazar & David T. Neal, *Habits and Goals in Human Behavior: Separate but Interacting Systems*, 17 PERSP. PSYCH. SCI. 590, 591 (2021) ("Once the response is triggered in mind, people tend to act on it through *ideomotor* processes whereby the thought of a behavior brings about the corresponding physical response in a reflexive, automatic manner.").

[21] *See* B.R. Andrews, *Habit*, 14 AM. J. PSYCH. 121, 121 (1903) ("A habit, from the standpoint of psychology, is a more or less fixed way of thinking, willing, or feeling acquired through previous repetition of a mental experience."); WILLIAM JAMES, THE PRINCIPLES OF PSYCHOLOGY 24 (1914)("[W]e find ourselves automatically prompted to *think*, *feel*, or *do* what we have been before accustomed to think, feel, or do, under like circumstances, without any consciously formed *purpose*, or anticipation of results.") (quoting WILLIAM B. CARPENTER, PRINCIPLES OF MENTAL PHYSIOLOGY 344 (1874)); *cf.* Michael S. Moore & Heidi M. Hurd,

They might cause the actor to ignore certain facts or risks, for example.[22] They might affect the values the actor assigns to possible outcomes.[23] Or they might cause the actor unconsciously to exclude possible alternative courses of action from his or her deliberations.[24]

This is not to say, of course, that the criminal law shouldn't concern itself with time-slice rationality of the sort that is captured by the Hand formula. It makes sense for the criminal law to require, as one component of liability, that the risks posed by the actor's conduct (the risks that inhered in the circumstances known to him or her) actually were unjustified, whether the actor bothered to think about them or not.[25] That is why the criminal law requires wrongdoing. The significance of habit's pervasive role in human decision-making is, rather, merely that it shows how incomplete the wrongdoing analysis is—how much the wrongdoing component leaves out. It leaves out all the facts about how the actor actually arrived at a decision.

The facts about how the actor actually arrived at a decision are, presumably, the subject of the culpability requirement then. But how do these facts figure? On the prevailing view of the culpability requirement as a concession to human weakness, we presumably would view the actor's reliance on habit as an expression of human weakness.[26] On this view, the

---

Punishing the Awkward, the Stupid, the Weak, and the Selfish: The Culpability of Negligence, 5 CRIM. L. & PHIL. 147, 190 (2011) ("One does best . . . to develop heuristics, proxies, mottos, habits, and practices that allow one to make good decisions in circumstances that require hasty judgments, and that off-set, compensate for, or otherwise prevent the realization of one's own worst tendencies.").

[22] *See* WILLIAM JAMES, PSYCHOLOGY: THE BRIEFER COURSE 39 (Dover Publ'ns 2001) (1892) ("[W]hat is called our 'experience' is almost entirely determined by our habits of attention.").

[23] *C.f.* Craig Agule, *Distinctive Duress*, 177 PHIL. STUDIES 1007, 1008–09 (2020) (explaining how "finite, limited beings" preserve their "limited attention, limited cognitive power, and limited volitional power," and so improve their decision making over time, by making themselves "comparatively more sensitive to some reasons and comparatively less sensitive to others").

[24] BRATMAN, *supra* note 13, at 89 (discussing how "personal policies" (a category that includes at least some habits) will, together with "demands for consistency and coherence," cause the agent to rule out some alternative courses of action).

[25] *See* JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 171 (Clarendon Press 1907) (1823) (arguing that "punishment ought not to be inflicted . . . [w]here it is *groundless*: where there is no mischief for it to prevent; the act not being mischievous upon the whole").

[26] Perka v. The Queen, [1984] 2 S.C.R. 232, 248–49 (Can.). In *Perka*, the court explained the "residual defense of necessity" as a concession to "human weakness": "a liberal and humane criminal law cannot hold people to the strict obedience of laws in emergency situations where normal human instincts, whether of self-preservation or of altruism,

habits themselves would count as malign influences that prevented the actor from realizing that the conduct was wrongful. The actor's culpability would hinge on whether his or her vulnerability to these malign influences represented understandable human weakness, or did not.

The central difficulty with this view is its assumption that the actor's habits represent malign influences on the actor's deliberative process. This assumption seems plausible enough at first glance. After all, if the actor's habit caused him or her to engage in wrongdoing, then the habit must itself be irrational, right?

Actually, no. As I will argue, even rational habits—that is, even habits that over an extended period will increase the frequency with which the actor does the *right* thing—sometimes will result in conduct that is wrongful from a time-slice perspective.[27] What is more, when a rational habit generates conduct that is wrongful or irrational from a time-slice perspective, the conduct itself will, paradoxically, partake of the habit's rationality. So the conduct will be irrational from one perspective and rational from another perspective. The two perspectives cannot be combined, moreover, or collapsed into one another. Rationality is irreducibly twofold.

Which explains why criminal liability, too, is twofold.[28] Criminal law's wrongdoing component measures the conduct's rationality from a time-slice perspective. The culpability component, by contrast, measures the conduct's rationality from a temporally extended perspective. This understanding of culpability, as compared with the dominant "concession to human weakness" understanding, doesn't just have the advantage of capturing the twofold nature of rationality. It also has the important practical advantage of being meaningful. For the empty question whether the factfinder can identify with the actor, it substitutes a question no less concrete than the question posed by the Hand formula itself, namely, whether the risks posed by the habits that generated the conduct outweigh the habits' benefits.

---

overwhelmingly impel disobedience." *Id.* at 249. For a different view of emergencies, see Paul J. Heald, *Mindlessness and Nondurable Precautions*, 27 GA. L. REV. 673, 688 (1993), where the author argues that "[w]hen accidents are caused by the instinctive invocation of a precautionary script, [tort] liability may sometimes be excused," but only if the "script" or habit is "desirable."

[27] *See* DANIEL KAHNEMAN, THINKING FAST AND SLOW 416 (2011) ("System 1 is indeed the origin of much that we do wrong, but it is also the origin of most of what we do right . . . .").

[28] *Cf.* Heidi M. Hurd & Michael S. Moore, *The Ethical Implications of Proportioning Punishment to Deontological Desert*, 15 CRIM. L. & PHIL. 495, 497 (2021) ("Moral desert consists of two distinct although related properties, wrongdoing and culpability.").

This Article will be divided into three parts. Part I will focus narrowly on behavioral habits—on habits that tell the actor what to *do*, rather than what to think or feel. Within this narrow framework, I will argue that even desirable habits can generate wrongful conduct and that, when they do, punishment of the actor for the habit-generated conduct is inconsistent with the criminal law's aims. I will also suggest, more ambitiously, that the distinctive kind of rationality associated with desirable habits is what the culpability component of criminal liability really is all about.

Part II will expand the argument's scope. As William James said, habits influence not only what we do but what we *think* and *feel* as well.[29] In Part II, I will argue that desirable habits of thought and feeling, like desirable habits of doing, can generate wrongful conduct. And I will argue that the exculpatory import of these desirable habits of thought and feeling explains traditional culpability doctrines. The best explanation for the defense of duress, for example, is that the defense evolved to exculpate actors whose conduct was generated by desirable habits of loyalty. Likewise, the best explanation for the heat-of-passion defense is that the defense evolved to exculpate in part actors whose homicidal anger was generated by desirable habits of pride.

Part III will briefly explore two alternative explanations for the seeming relationship between the desirability of an actor's habits and the actor's culpability. First, I will consider the possibility that, although habits sometimes negate culpability, this culpability-negating effect does not depend on the habit's desirability. Second, I will consider the possibility that the actor's habits have only evidentiary significance. What really matters, on this latter view, is the actor's moral character.

I. JUDGING THE CULPABILITY OF HABITUAL BEHAVIOR

Behavior that satisfies the criminal law's wrongdoing component usually is conscious rather than purely habitual. But what happens when wrongful conduct *is* habitual—when the actor's habits don't just influence his or her deliberations but, rather, generate the wrongful conduct directly? How do we judge the actor's culpability in these cases? The answer, as I will argue in this Part, is that the actor's culpability depends on the desirability of the habits that generated the actor's wrongful conduct.

---

[29] JAMES, *supra* note 21, at 24.

A. FRAMING THE QUESTION: *STATE V. LEWIS*

Consider *State v. Lewis*.[30] Devin Lewis was driving a tractor-trailer rig westbound on Highway 18 in Oregon when the vehicle traveling in front of him, a catering truck, slowed and then stopped in preparation for making a left turn.[31] Two logging trucks were approaching the intersection from the other direction, so the driver of the catering truck had to wait for them to pass.[32] Lewis, who was behind the catering truck, apparently did not notice at first that it had signaled a turn.[33] When he did notice, too late, he slammed on his truck's brakes, putting it into a skid.[34] As Lewis's truck skidded forward, the rear end of his empty trailer clipped the rear of the catering truck, knocking the catering truck into the path of an oncoming logging truck.[35] The resulting collision killed the driver of the catering truck.[36] Lewis later was indicted for criminally negligent homicide for causing the other driver's death.[37]

Three features of Lewis's case make it useful as a vehicle for exploring the relationship between culpability and habit, as I will explain in this section. First, Lewis's conduct appears to be objectively wrongful. Which means that his liability ultimately will hinge on the question whether his conduct also was culpable. Second, Lewis's wrongful conduct, his inattention to the roadway in the moment of the accident, was likely a product of habit, rather than of deliberation. Third, there is a real question, given the circumstances, whether Lewis's conduct really was culpable.

Let's start with the first of these three features—the conduct's wrongfulness. In the criminal law, wrongdoing and culpability traditionally have been treated as "conceptually distinct" components of liability.[38] In

---

[30] 290 P.3d 288 (Or. 2012).

[31] *Id.* at 289.

[32] *Id.*

[33] *Id.*

[34] *Id.* at 289–90.

[35] *Id.* at 290.

[36] *Id.* at 289.

[37] *Id.*

[38] Anthony M. Dillof, *Punishing Bias: An Examination of the Theoretical Foundations of Bias Crime Statutes*, 91 NW. U. L. REV. 1015, 1029 (1997) ("[U]nder both objective and subjective theories, wrongdoing and culpability are conceptually distinct components for determining desert."); *see also* George P. Fletcher, *The Theory of Criminal Negligence: A Comparative Analysis*, 119 U. PA. L. REV. 401, 427 (1971) ("[T]here is a distinction . . . between the legality of conduct and the culpability of the individual who engages in the conduct."); Heidi M. Hurd, *The Deontology of Negligence*, 76 B.U. L. REV.

criminal negligence cases like Lewis's, for example, this distinction is reflected in the "two distinct functions" assigned to the factfinder.[39] In jurisdictions that have adopted the Model Penal Code's influential definition of criminal negligence, as Oregon has,[40] the factfinder's first responsibility is to decide whether the actor's conduct was *wrongful*—that is, whether the risk posed by the conduct was both substantial and unjustifiable.[41] The factfinder's second responsibility, naturally, is to "make the *culpability* judgment."[42]

The Model Penal Code is famously unhelpful in explaining how the jury is to make the culpability judgment. The Code commentary says: "[I]t is quite impossible to avoid tautological articulation of the final question. The tribunal must evaluate the actor's failure of perception and determine whether, under all the circumstances, it was serious enough to be condemned."[43] By contrast, the factfinder's responsibilities in relation to the wrongdoing component are clear: The factfinder must evaluate the risk posed by the actor's conduct under "the circumstances known to him,"[44] then must decide whether this risk is "unjustifiable," given the countervailing benefits of the conduct.[45] Where wrongfulness is concerned, then, the Model Penal Code just adopts a kind of criminal law variant of the Learned Hand formula.[46]

It is reasonable to suppose that Lewis's conduct at least was wrongful, regardless of whether it also was culpable. It is reasonable to suppose, in other words, that the risk posed by Lewis's inattention to the road ahead in

---

249, 262 (1996) ("Generally speaking, defendants must be found *both* to have done a bad act and to have done it culpably before they can properly be held liable or punished.").

[39] MPC COMMENTARIES, *supra* note 10, at 241.

[40] OR. REV. STAT. § 161.085(10) (1972). The Model Penal Code's definitions of recklessness and criminal negligence are among the most influential aspects of the Code. *See* 1 WAYNE R. LAFAVE, SUBSTANTIVE CRIMINAL LAW § 5.4(b) at 501–02 (3d ed. 2017) (observing that "in most recent recodifications the Model Penal Code approach [to defining recklessness and negligence] has been substantially followed"). Though Oregon's definition of criminal negligence varies the language of Model Penal Code § 2.02(2)(d) somewhat, the core elements—the requirement that the conduct pose a "substantial and unjustifiable risk" and the requirement of a "a gross deviation from the standard of care" —are unchanged.

[41] *See* MPC COMMENTARIES, *supra* note 10, at 241 (explaining that the first of the jury's two functions in negligence cases "is to examine the risk and the factors that are relevant to its substantiality and justifiability").

[42] *Id.* (emphasis added).

[43] *Id.*

[44] MODEL PENAL CODE § 2.02(d) (AM. L. INST. 1985).

[45] *Id.*

[46] Dressler, *supra* note 15, at 957.

the moment before the accident was not offset by the inattention's benefits. The trial court emphasized some of the facts that enhanced the risk in Lewis's case: "Defendant's trailer was empty, making it difficult to stop, and defendant knew that fact." "The highway was wet." "The highway had only two lanes, which meant defendant was driving under circumstances where he knew that he had less room to maneuver."[47] Even in the absence of these circumstances, though, Lewis's inattention probably would have been wrongful. The criminal law's Hand-formula test for wrongdoing is basically the same as the Hand-formula test for civil negligence.[48] And courts generally agree that rear-ending another car or truck qualifies as civil negligence, at least in the absence of extenuating circumstances.[49] Indeed, Lewis appears to have conceded in the Oregon Supreme Court that his conduct was at least civilly negligent.[50]

A second important feature of Lewis's case is that Lewis's conduct probably was habitual, rather than deliberative. The conduct that provides the most plausible locus for Lewis's prosecution is his "act," such as it is, of directing his attention away from the roadway in the moment before the crash.[51] Decisions about how to allocate one's attention while driving usually are habitual, as for that matter are most decisions about how to allocate one's attention.[52] The human mind has "a limited stock of attention resources."[53]

---

[47] State v. Lewis, 290 P.3d 288, 291 (Or. 2012).

[48] Dressler, *supra* note 15, at 957; *see also* Anthony M. Dillof, *Transferred Intent: An Inquiry into the Nature of Criminal Culpability*, 1 BUFF. CRIM. L. REV. 501, 534 (1998) ("Tort and criminal law should be understood as pursuing different visions of justice in response to the same underlying phenomena: wrongdoing.").

[49] *See, e.g.*, Carter v. Williams, 361 F.2d 189, 193–94 (7th Cir. 1966) ("[I]t is the general rule in Illinois that it is negligence as a matter of law to drive an automobile at such a rate of speed that it cannot be stopped in time to avoid a collision with an object discernible within the driver's length of vision ahead of him."); Davis v. Quinones, 743 N.Y.S.2d 171, 172 (App. Div. 2002) ("A rear-end collision with a stopped automobile creates a prima facie case of negligence with respect to the operator of the moving vehicle and imposes a duty on that operator to provide a nonnegligent explanation as to how the accident occurred.") (citation omitted).

[50] *Lewis*, 290 P.3d at 297 (attributing to defendant Lewis the view that the accident was "a case of mere negligence").

[51] *See* Alan C. Michaels, *Constitutional Innocence*, 112 HARV. L. REV. 828, 877–78 (1999) (explaining that "punishment must be predicated on some voluntary act or omission").

[52] *Cf.* DUHIGG, *supra* note 18, at 17 (explaining habit's role in "backing your car out of the driveway").

[53] CHRISTOPHER CHABRIS & DANIEL SIMONS, THE INVISIBLE GORILLA 24 (2010); KAHNEMAN, *supra* note 27, at 23 ("The often-used phrase 'pay attention' is apt: you dispose of a limited budget of attention that you can allocate to activities, and if you try to go beyond your budget, you will fail.").

Habit plays a pervasive role in the mind's allocation of these resources, as William James observed in his groundbreaking account of the stream of consciousness: "[W]hat is called our 'experience' is almost entirely determined by our habits of attention."[54] It is conceivable, no doubt, that Lewis made a conscious decision to direct his attention away from the roadway in the moment before he noticed the catering truck. It seems more likely, though, that this decision, like most other "decisions" about how to allocate one's attention while driving, was automatic.[55]

A third helpful feature of Lewis's case is that, realistically, the question of culpability might be resolved either way. As a preliminary matter, the mere fact that Lewis's inattention was a product of habit rather than of deliberation does not, in itself, either absolve him of culpability or relieve the government of the burden of proving culpability. Though criminal law scholars occasionally have expressed concern about how to justify criminal liability for habitual conduct, nobody really questions the traditional view that habitual acts may trigger criminal liability.[56] The Model Penal Code, for example, defines "voluntary act" to exclude any "bodily movement that otherwise is not a product of the effort or determination of the actor, either conscious *or habitual*."[57] So the fact that conduct is habitual obviously does not foreclose the assignment of liability. Nor, of course, does the fact that the conduct is habitual somehow relieve the government of the burden of proving culpability.[58] In cases where the wrongful act is habitual, then, as in cases

---

[54] JAMES, *supra* note 22, at 39.

[55] *See* Michael S. Moore, *Addiction, Responsibility, and Neuroscience*, 2020 U. ILL. L. REV. 375, 399 (2020) ("With skills that we have mastered, or habits that we have acquired—such as, for example, going to lunch at our regular time, driving a curvy mountain road, stacking lumber, and playing the piano—we can do such things without consciously directing the routines by which we do them.").

[56] *See* Moore & Hurd, *supra* note 21, at 154 ("We think that some instances of dispositional awareness/belief can constitute advertence. This kind of awareness/belief is all any actor ever has as he exercises well-honed habits and skills, such as driving an automobile or ironing clothes or mowing a lawn or executing a repetitive task on an assembly line. Yet we are responsible for the way that we do these activities."); Gideon Yaffe, *The Voluntary Act Requirement* in THE ROUTLEDGE COMPANION TO THE PHILOSOPHY OF LAW 174, 176 (Andrei Marmor ed., 2012) ("If a defendant is . . . shown to have engaged in a habitual bodily movement, then that's good enough to comply with the restriction imposed by the [voluntary act requirement]."). *But see* Kimberly Kessler Ferzan, *Opaque Recklessness*, 91 J. CRIM. L. & CRIMINOLOGY 597, 646–47 (2001) (raising concerns about the punishment of habitual actions and discussing possible rationales for such punishment).

[57] MODEL PENAL CODE § 2.01(2)(d) (AM. L. INST. 1985) (emphasis added).

[58] *See* Eric A. Johnson, *The Crime That Wasn't There: Wyoming's Elusive Second-Degree Murder Statute*, 7 WYO. L. REV. 1, 19–20 (2007) ("A voluntary act, though necessary to justify criminal liability, is not close to being sufficient.").

where the wrongful act is conscious, the government still must prove that the actor's conduct was culpable.

From a factual perspective, too, the question of culpability in Lewis's case seems as though it might be resolved either way. In cases like Lewis's, where a motorist causes a fatal accident by rear-ending another vehicle whose driver has stopped or slowed to make a left turn, courts sometimes conclude that the motorist's conduct is culpable and sometimes do not.[59] The results depend on the facts.[60]

B. LAPSE CASES AND THE INSIGHTS OF TORT SCHOLARS

In cases like Lewis's, where the actor's wrongful conduct was generated by habit rather than by a conscious exercise of volition, how do we go about deciding whether the actor's wrongdoing also was culpable?

Courts and scholars have taken a variety of approaches to this question. For example, some would argue that the actor's culpability depends on whether he or she had the "capacity" in the moment of the wrongful act to avoid committing the wrongful act.[61] Others would argue instead that the actor's culpability depends simply on the degree of disproportion between the risks posed by the actor's act and the act's utility—on whether the actor's conduct "involves greater risk of harm to others, without any compensating social utility, than does simple negligence."[62]

---

[59] *Compare* State v. Lewis, 290 P.3d 288, 299 (Or. 2012), where the court upheld the defendant's conviction for criminally negligent homicide, *with* Queeman v. State, 520 S.W.3d 616, 619 (Tex. Crim. App. 2017), where, on strikingly similar facts, the court concluded that evidence was "legally insufficient to establish criminally negligent homicide."

[60] *See, e.g.*, Lewis, 290 P.3d at 296 ("We agree that, if this case involved only a driver's brief inattention to his driving and nothing more, the evidence likely would fall short of establishing criminal negligence.").

[61] DAVID O. BRINK, FAIR OPPORTUNITY AND RESPONSIBILITY 158 (2021) (identifying "broad culpability" with "two main conditions: the agent's internal capacities or normative competence and her opportunities or situational control"); H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 149–52 (1968) (applying capacity analysis to criminal liability for negligence).

[62] Queeman v. State, 520 S.W.3d 616, 623 (Tex. Crim. App. 2017) ("[C]onduct that constitutes criminal negligence involves a greater risk of harm to others, without any compensating social utility, than does simple negligence (citing Montgomery, 369 S.W.3d at 193)); *see also* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 127 & n.89 (8th ed. 2018) (suggesting that what distinguishes criminal negligence from civil negligence is the degree of disproportionality between the risk and the utilities: "Applying the Learned Hand formula, *criminal* negligence exists when 'PL' *far* outweighs 'B'; the level of negligence should be so great 'that it would be shocking to allow the actor's lack of awareness to excuse his actions in the circumstances.'").

I will take a different approach, as I have said. I will argue that in cases like Lewis's, the actor's culpability depends at least in part on the desirability of the habits that generated the actor's conduct. The actor's culpability depends at least in part, that is, on whether the habits that generated the actor's conduct were of a kind whose internalization by the actor was likely over time to produce more benefits than costs. To make this claim work, I will first need to show, somewhat counterintuitively, that even desirable habits can generate wrongful conduct. After all, if habits that generate wrongful conduct always are *undesirable*, then the separate culpability inquiry could not very well hinge on whether the habits that generated the actor's wrongful conduct were desirable.

In what sorts of cases, then, can desirable habits generate wrongful conduct? To begin with, desirable habits can generate wrongful conduct in cases where, paradoxically, occasional lapses are beneficial—where occasionally doing the *wrong* thing improves the actor's overall performance over time. These cases, which I will call "lapse" cases, share three basic features: (1) over an extended sequence of moments, the actor faces a series of similar or identical choices, A_1, A_2, . . . A_n; (2) in each of these individual moments, the balance of risks and benefits favors behavior *V*; and (3) still, everyone will be better off if the actor occasionally chooses *not-V*.

Consider Lewis's case. We have already stipulated, for the sake of discussion, that Lewis's inattention in the moment before the accident was wrongful. We have stipulated, in other words, that the risk posed by Lewis's inattention to the road in the moment before the accident was not offset by the inattention's utility. It seems plausible to suppose, too, that the same thing was true of every other particular moment on Lewis's journey. It seems plausible to suppose, that is, that at any particular moment in Lewis's journey, the best use of Lewis's "limited stock of attention resources" was, on balance, to focus intently on the road ahead.

If this conclusion seems plausible given how the criminal law defines wrongdoing, it also seems troubling. After all, even reasonably careful drivers occasionally direct their attention away from the road ahead. It would be too taxing, and perhaps even counterproductive, to try to maintain a constant, laser-like focus on the roadway throughout every moment of an extended journey.[63] Tort scholar Lee Fennell has made this very point in connection with tort negligence:

---

[63] Robert L. Rabin, *The Fault of Not Knowing: A Comment*, 4 THEORETICAL INQ. L. 427, 432 (2003) ("When engaging in repetitive, risk-related activities, it is impossible for *anyone* to achieve a perfect compliance rate—that is, to achieve a constant state of perfect

> It is impossible for human beings to be perfectly consistent in taking precautions that must be repeated over and over in real time, such as alertly scanning the road while driving. A single moment of inattention that produces an accident might be part of a larger pattern that represents as much care as any person could reasonably be expected to exercise.[64]

The Restatement (Third) of Torts acknowledges the same basic point, namely, that wrongful—or negligent—conduct might be part of a larger pattern of behavior that, in the aggregate, is entirely reasonable.[65] In the Restatement example, a factfinder faces the question "whether a pedestrian should have detected, and hence avoided stumbling over, a banana peel located on a crowded sidewalk."[66] The Restatement's authors acknowledge that even an entirely reasonable pedestrian might notice the banana peel only "nine times out of ten."[67] In other words: Though in any *particular* moment the balance of risks favors keeping an eye on the sidewalk, even an entirely reasonable pedestrian might occasionally divert his attention from the sidewalk. If this entirely reasonable pedestrian winds up stumbling on a banana peel in a moment of inattention, however, the pedestrian's conduct still will qualify as negligent, according to the Restatement.[68] Negligence, after all, depends on the risks and benefits of a *particular* act or omission.[69] It does not depend on patterns of behavior that are reasonable "over an extended period of time," say the Restatement's authors.[70]

---

attentiveness and performance."); *cf.* R.M. HARE, MORAL THINKING: ITS LEVELS, METHOD, AND POINT 200–01 (1981) (exploring generally the possibility that actors can end up "getting themselves and others into serious trouble" by "demand[ing] too much of themselves").

[64] Lee Anne Fennell, *Accidents and Aggregates*, 59 WM. & MARY L. REV. 2371, 2374 (2018) [hereinafter Fennell, *Accidents and Aggregates*]; *see also* Mark F. Grady, *Efficient Negligence*, 87 GEO. L.J. 397, 401 (1998) ("[I]t is impossible for real-world actors to achieve this standard of perfection during every waking moment; they will lapse. If some of these lapses are inevitable, it is reasonable to conclude that even more of them are efficient. Typically, an efficiency standard does not require people to do everything possible to avoid harmful outcomes. Moreover, it seems highly probable that increasing amounts of perfection come at increasing marginal costs.").

[65] *See* RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL AND EMOTIONAL HARM § 3 cmt. k at 36–37 (2010).

[66] *Id.* at 37.

[67] *Id.*

[68] *Id.*

[69] *Id.* § 3 cmt. e at 30–31.

[70] *Id.* § 3 cmt. k at 37; *see also* Fennell, *supra* note 64, at 2374 ("Alternatively, the accident-causing shortfall might be a representative draw from an urn of chronically unreasonable conduct. Tort law does not distinguish between these cases because it takes as the relevant unit of analysis the single accident-causing moment, not a larger behavioral

Cases of this kind arise from what some economists call a "scale mismatch."[71] As Professor Fennell has explained, a scale mismatch occurs "where the relevant benefit (or cost) of a decision"—of a decision to direct one's attention away from the sidewalk, for example—is experienced only at an aggregate scale, while the corresponding inputs must be made at a decision-by-decision scale."[72] When we apply the Hand formula, we judge the rationality of a particular act or omission at a decision-by-decision or "time-slice" scale.[73] That is, we concern ourselves exclusively with the downstream risks and utilities of the particular act or omission as calculated in that moment.[74] By comparison, when we say that a driver or pedestrian will perform better over time if they occasionally permit their attention to lapse, we consider the actor's behavior in the aggregate; we take a "temporally extended" perspective on the conduct, in philosopher Michael Bratman's phrase.[75]

At first glance, it might seem as though it would be possible for the time-slice Hand formula to capture somehow the larger-scale efficiencies associated with occasional departures or "lapses."[76] As lawyers know, the Hand formula can capture lots of indirect costs and benefits. It can capture the conduct's long-term effects on social norms, for example.[77] As it turns out, though, the time-slice Hand formula cannot capture the efficiencies associated with occasional departures or lapses. The trouble isn't that the costs and benefits associated with occasional lapses "are spread out over

---

sample that might provide corroborating or mitigating evidence about the actor's overall level of care.").

[71] Drazen Prelec, *Limitations on Traditional Economic Analysis* in SOCIO-ECONOMICS: TOWARD A NEW SYNTHESIS 131, 134 (Amitai Etzioni & Paul R. Lawrence, eds. 1991) (explaining that a "scale mismatch" occurs when "one element of the cost-benefit pair is perceived to have impact only in an aggregate sense, that is, only if the same action is repeated many times, or on a larger scale").

[72] Lee Anne Fennell, *Personalizing Precommitment*, 86 U. CHI. L. REV. 433, 433–34 (2019).

[73] BRATMAN, *supra* note 13, at 78.

[74] *See* RESTATEMENT (THIRD) OF TORTS, *supra* note 65, § 3 cmt. e at 30–31.

[75] BRATMAN, *supra* note 13, at 78.

[76] Tort scholars refer to these departures from time-slice rationality pejoratively as "lapses," despite acknowledging that these lapses may be efficient over the long term. *See* Grady, *supra* note 64, at 400–01; Fennell, *Accidents and Aggregates*, *supra* note 64, at 2374; Christopher Brett Jaeger, *The Empirical Reasonable Person*, 72 ALA. L. REV. 887, 903 (2021).

[77] *See* Stuart P. Green, *Looting, Law, and Lawlessness*, 81 TUL. L. REV. 1129, 1152–53 (2007) (discussing how indirect harms from looting—loss of civil order, sense of fear, "the possibility that looting tends to be contagious"—would affect the balancing of evils in application of the necessity defense to looting).

time."[78] Rather, the trouble is that the opportunities to lapse are spread out over time. If they were not—if the actor could realize the gains associated with lapsing only by lapsing in a particular moment—then in that moment the balance of risks and benefits would favor lapsing. At least where unstructured activities like driving are concerned, however, one moment is as good as any other for lapsing.[79] In any particular moment, then, the balance of risks and benefits always favors *not* lapsing.

Of course, the fact that these benefits cannot be captured by the Hand formula doesn't make them any less real, or any less deserving of the attention of rational actors. An actor who ignored the benefits of occasional departures or lapses—who ignored the demands of temporally extended rationality in favor of always relying on a time-slice Hand calculation—might wind up engaging in "deeply suboptimal" conduct, as Professor Fennell has said.[80] Nor would we be justified in concluding—from the fact that the Hand formula cannot capture the long-term benefits of occasional departures—that we cannot evaluate the rationality of these occasional departures. We can, at least where the departures are generated by a habit.[81] In conducting this evaluation, we still would balance risks and benefits, just as we do when we judge the rationality of the actor's conduct in a particular moment under the Hand formula.[82] But the risks and benefits we would balance are the risks and benefits of developing particular habits, not of engaging in a particular act.[83]

---

[78] George Loewenstein & Richard H. Thaler, *Anomalies: Intertemporal Choice*, 3 J. ECON. PERSP. 181, 181 (1989) ("Intertemporal choices, decisions in which the timing of costs and benefits are spread out over time, are both common and important.").

[79] *Cf.* DAN B. DOBBS, PAUL T. HAYDEN & ELLEN M. BUBLICK, HORNBOOK ON TORTS 278 (2d ed. 2016) (acknowledging that the existence of "designs, plans, or systems for dealing with hazards" potentially would make the actor's rate of lapsing subject to risk-utility balancing).

[80] Fennell, *Personalizing Precommitment*, *supra* note 72, at 447.

[81] For an explanation of why this sort of analysis might be limited to cases where the departures are generated by a habit or a general, *see* discussion *infra* Part I(d).

[82] *See* BRATMAN, *supra* note 13, at 52 (observing that we can assess "habits and patterns in terms of their long-range consequences of their internalization by limited agents like us"); Grady, *supra* note 64, at 400–01 ("Using the Learned Hand formula, a court might decide that drivers should look for pedestrians twenty times per minute. Exactly how often drivers must look, a court might reason, would depend upon Learned Hand type factors."); Heald, *supra* note 26, at 688 (assuming that courts could evaluate the desirability of particular psychological "scripts").

[83] *Cf.* Hooker, *supra* note 18 (summarizing the view that "the best kind of rule-consequentialism ranks systems of rules *not* in terms of the expected good of *complying* with them, but in terms of the expected good of their *acceptance*").

In Lewis's case, for example, it would be possible to evaluate in just this way the habits of attention that generated the lapse in Lewis's attention just before the crash.[84] This calculation would require us to consider, for example, the frequency of these habitual lapses. After all, as the frequency of lapses increases, so does the likelihood that a lapse will coincide temporally with one of those rare circumstances where even a short lapse in attention is sufficient to cause an accident. The evaluation of Lewis's habits also would require us to consider the diminishing returns of occasional lapses as the lapses become more frequent. By considering factors like these, we could decide whether the habit that generated the lapse was rational or desirable.

With the help of tort scholars, then, we appear to have succeeded in identifying a class of cases where desirable habits can generate wrongful conduct, namely, cases where an occasional "lapse" improves the actor's performance over an extended period. In a moment, we will turn to the criminal law implications of these cases. Before we do, though, let's discuss another, much broader class of cases where even desirable habits can generate wrongful conduct, namely, cases where habits are over-inclusive.

C. OVER-INCLUSIVENESS CASES

In the previous Section, we assumed, not implausibly, that a driver who occasionally diverts his or her attention from the roadway may, as a result, perform better over time. We assumed, in other words, that occasional moments of inattention might produce benefits over time in the form of increased attention in *other* moments. In this Section, let's ignore the apparent benefits of occasional lapses. Let us assume, for the sake of argument, that over time a driver is better off just paying as much attention in any particular moment as that moment demands. Of course, these demands will vary from moment to moment. In some moments, the balance of risks and benefits might require the driver to devote 40% of his attention to the road ahead. In other moments, it might require 70%. In any case, in this new picture, inattention in one moment does not result directly in improved attention in other moments.

Is it still possible, in this new picture, for desirable habits to generate wrongful conduct? Suppose, for example, that in the moment before the accident, the balance of risks and utilities required Lewis to devote 70% of his attention to the roadway: the road was wet, he was approaching a junction with another road, etc.[85] In that moment, however, Lewis was devoting only

---

[84] *See* Grady, *supra* note 64, at 400–01.

[85] State v. Lewis, 290 P.3d 288, 291 (Or. 2012).

50% of his attention to the roadway. Suppose, however, that Lewis's level of attention in this moment was not the product of deliberation, but, rather, was the product of a habit that generates a 50% level of attention across a wide range of circumstances, in most of which the balance of risks and benefits would demand only 30% of Lewis's attention. Suppose, that is, that Lewis's relative inattention was a product of a habit that proved *over-inclusive* in this particular case.[86] Is it possible, in this scenario, that Lewis's relative inattention in the critical moment, despite being wrongful, still was the product of a desirable habit—still was the product of temporally extended rational agency?

It is, I think. If the habit's costs in over-inclusiveness were counterbalanced by the kinds of benefits usually associated with habits, Lewis's wrongful conduct still might be rational from a temporally extended perspective. This should not sound surprising, at least to lawyers. Habits are a species of rule, after all.[87] And as every lawyer knows, even good rules inevitably are overinclusive.[88] When a rule's over-inclusiveness is justified by the rule's long-term benefits, moreover, an application of the rule will not usually be subject to criticism merely on the ground that the rule was overinclusive as applied in that particular instance.[89]

---

[86] For our purposes, a habit is overinclusive if (1) the habit causes the actor to react with behavior *V* across a whole range of similar but not identical situations, A_1, A_2, . . . A_n, and (2) in one or more of these specific situations, a case-specific balancing of risks and benefits would favor *not-V*. *See* FREDERICK SCHAUER, PLAYING BY THE RULES 50 (1991) (explaining that a rule is overinclusive in relation to its generating justification where the rule covers situations "that would not be covered by a direct application of the rule's justifications").

[87] *See* Geoffrey M. Hodgson, *The Ubiquity of Habits and Rules*, 21 CAMBRIDGE J. ECON. 663, 664 (1997) (observing that rules and habits obviously "have the same general form: in circumstances *X*, action *Y* follows").

[88] Andrew Coan & Nicholas Bullard, *Judicial Capacity and Executive Power*, 102 VA. L. REV. 765, 812 (2016) (observing that categorical rules "are inevitably both over- and under-inclusive relative to their underlying purpose"); Neil Kinkopf, *Of Devolution, Privatization, and Globalization: Separation of Powers Limits on Congressional Authority to Assign Federal Power to Non-Federal Actors*, 50 RUTGERS L. REV. 331, 350 (1998) ("Categorical rules are inevitably over- or under-inclusive, or both."); David Lyons, *Utility and Rights*, NOMOS XXIV: ETHICS, ECONOMICS, AND THE LAW,107, 123 (J. Roland Pennock & John W. Chapman eds., 1982) (Yearbook of the American Society for Political and Legal Philosophy) ("[I]t is predictable that real social rules that are supported by the best utilitarian and economic arguments will require decisions in particular cases that would not most effectively promote welfare or efficiency."); SCHAUER, *supra* note 86, at 135 ("Rule-based decision-making . . . entails an inevitable under- and over-inclusiveness . . . .").

[89] *See, e.g.*, New York City Transit Auth. v. Beazer, 440 U.S. 568, 592 (1979) (rejecting Equal Protection challenge to Transit Authority rule which excluded from employment any former heroin addicts who were currently receiving methadone treatment, despite

To illustrate, let's consider a standard-issue example of a legal rule: Federal Evidence Rule (FRE) 404, which excludes most character evidence.[90] Suppose a criminal defendant's trial strategy is to shift blame to another suspect. To this end, she wants to introduce evidence of the other suspect's prior convictions for similar offenses. The judge rules, correctly, the other suspect's prior convictions are barred by the character rule, FRE 404.[91] Now consider: Under what circumstances would the exclusion of the defendant's evidence under the rule fairly be subject to criticism?

We could, in principle, evaluate the judge's ruling purely in terms of the likely downstream consequences, good and bad, of excluding this particular piece of evidence in this particular case. Roughly, this sort of balancing is what's contemplated by FRE 403, which requires the judge to balance the evidence's probative value—its value to the jury in resolving the dispute—against its tendency to create prejudice against one or another party, or to distract or confuse the jury.[92]

This is not the way we would evaluate the judge's ruling, however. Because the character rules unambiguously exclude this evidence, the judge had no occasion to consider how the evidence would affect the jury's verdict.[93] The judge just applied a rule, as judges often do. To evaluate the

---

acknowledging that the rule might well be substantially overinclusive); Holmes v. South Carolina, 547 U.S. 319, 320 (2006) (explaining that exclusion of defense evidence under a categorical evidence rule will violate the accused's constitutional rights only where the rule's costs in over-inclusiveness are "disproportionate to the ends that [the rule is] asserted to promote"); *cf.* SCHAUER, *supra* note 86, at 149 (acknowledging the efficiencies associated with rule-following are "often are sufficient to warrant tolerating some number of suboptimal results").

[90] FED. R. EVID. 404(a).

[91] *See* Hauk v. Indiana, 729 N.E.2d 994, 1002 (Ind. 2000) (holding that evidence of co-defendant's character, when offered to prove that the co-defendant had committed the murder without the defendant's willing participation, was properly excluded under Indiana rule closely modeled on FRE 404).

[92] *See* Edward J. Imwinkelried, *The Meaning of Probative Value and Prejudice in Federal Rule of Evidence 403: Can Rule 403 Be Used to Resurrect the Common Law of Evidence?*, 41 VAND. L. REV. 879, 884, 889–93 (1988) (discussing balancing the probative value and prejudicial impact under FRE 403 and addressing, in addition, the question whether the judge has the power under FRE 403 to exclude evidence "in pursuit of an extrinsic social policy").

[93] *See* DAVID P. LEONARD, EDWARD J. IMWINKELRIED, DAVID H. KAYE, DAVID E. BERNSTEIN, ANDREW FERGUSON, JENNIFER L. MNOOKIN, MAGGIE WITTLIN, ROGER PARK & TOM LININGER, THE NEW WIGMORE: A TREATISE ON EVIDENCE § 4.10 (2017) ("Some evidence rules or parts of rules do not give trial courts any discretion. For example, Federal Rule 404(a) provides that '[e]vidence of a person's character or a trait of character is not admissible for the purpose of proving action in conformity therewith on a particular occasion . . . .'").

judge's ruling as if the judge instead had balanced the costs and benefits of excluding the evidence would be misconceived.

Does that mean the judge's application of the character rule, if correct, is exempt from criticism? No. To criticize the ruling, though, we would have to attack the categorical rule on which the judge's ruling was based. We would have to argue that the categorical rule itself was irrational—that the costs of the rule are "disproportionate to the ends that [the rule is] asserted to promote."[94] To show that the rule itself was irrational, moreover, it wouldn't be enough to show that the rule was over-inclusive as applied in a particular case.[95] It wouldn't be enough to show that the costs of excluding the evidence in this case exceeded the benefits. After all, categorical rules "are inevitably both over- and under-inclusive relative to their underlying purpose."[96] The question in cases like this is whether the rule's costs—including the costs associated with over-inclusiveness—are offset by the rule's efficiencies.[97]

They often are. For starters, categorical rules save time and resources by enabling the decision-maker to forego careful, case-specific balancing of costs and benefits.[98] Moreover, in spite of their inevitable over-inclusiveness, categorical rules may produce a larger number of correct decisions across the universe of cases than would a case-specific balancing, especially in

---

[94] Holmes v. South Carolina, 547 U.S. 319, 324 (2006) (quoting United States v. Scheffer, 523 U. S. 303, 308 (1998).

[95] *See id.* at 325 (describing as arbitrary "rules that exclude[] important defense evidence" without serving an legitimate interests); *cf.* Joe Mintoff, *Can Utilitarianism Justify Legal Rights with Moral Force?*, 151 U. PA. L. REV. 887, 890 (2003) ("[T]he legal institutions which maximally promote human welfare will inevitably permit (if not require) actions which do not maximally promote human welfare.").

[96] Andrew Coan & Nicholas Bullard, *Judicial Capacity and Executive Power*, 102 VA. L. REV. 765, 812 (2016); *see also* Kinkopf, *supra* note 88, at 350 ("Categorical rules are inevitably over- or under-inclusive, or both.").

[97] Over-inclusiveness *is* a cost. Categorical evidence rules like FRE 404 ultimately are designed to promote the very interests—the ascertaining of truth, for example—that a decision maker would take into account in resolving cases on the basis of a case-specific balancing of risks and benefits. *See* Eleanor Swift, *One Hundred Years of Evidence Law Reform: Thayer's Triumph*, 88 CALIF. L. REV. 2437, 2446 (2000) ("Both probative value and risk of the Rule 403 dangers are taken into account in the drafting of the categorical requirements themselves."). When a categorical rule produces an outcome that wouldn't be justified under a case-specific balancing, this outcome is a cost.

[98] *See* Liesa L. Richter, *Posnerian Hearsay: Slaying the Discretion Dragon*, 67 FLA. L. REV. 1861, 1882 (2015) ("A likely casualty of adopting a generic reliability approach to hearsay would be litigation efficiency and valuation."); SCHAUER, *supra* note 86, at 145–49 (discussing "the argument [for rule-based decision-making] from efficiency").

situations where case-specific balancing is complex or difficult.[99] Sometimes case-specific balancing, far from producing careful, fine-tuned analyses of costs and benefits, instead produces decisions skewed by irrelevant factors or even by bias.[100] In summary, then, where a lower court judge correctly applies a categorical evidence rule, the judge's evidentiary ruling will be subject to criticism only if the rule itself is irrational—only if the rule's costs outweigh its benefits.

If rules can make conduct rational even when they are overinclusive, so can habits. The two central benefits of rule-following, increased efficiency and increased accuracy, both have their counterparts in habitual behavior. As William James said: "If practice did not make perfect, nor habit economize the expense of nervous and muscular energy, [the individual] would be in a sorry plight."[101] What James means by "practice . . . make[s] perfect" is that a musician, for example, is more likely to play the right notes after committing the piece to habit than before.[102] But habit also improves accuracy much as court rules do, by insulating the actor from influences that would otherwise lead to inaccurate decision making, for instance. As philosopher R.M. Hare has observed, decisions made in the moment of a dilemmatic choice may well be influenced unduly by our short-term interests, temptation, hot blood, etc.[103] Habits—habits of loyalty to friends, for example—insulate actors from these sorts of influences.[104] Moreover, as Bratman has said: "It may sometimes be easier to appreciate the expectable

---

[99] *See* Larry Alexander, *Introduction to the Symposium on the Rationality of Rule-Following*, 42 SAN DIEGO L. REV. 53, 54 (2005) (observing that "over the full range of cases to which they apply, rules may produce fewer moral errors than standards"); Richard A. Posner, *On Hearsay*, 84 FORDHAM L. REV. 1465, 1468 (2016) (observing that case-specific "rulings on objections to [hearsay] evidence would be difficult because often it's very difficult to estimate the reliability of a hearsay statement").

[100] *See* Richter, *supra* note 98, at 1886–87 (arguing that without a categorical hearsay rule, a judge's case-by-case reliability assessments would sometimes be influenced by the judge's perception of the "correct" outcome, particularly in "sympathetic, compelling, or politically charged cases"); SCHAUER, *supra* note 86, at 149–55 ("When free to take all relevant factors into account, decision-makers can err in numerous ways.").

[101] JAMES, *supra* note 22, at 5.

[102] *Id.* at 5–6 (identifying "false notes" as a pitfall of non-habitual (or pre-habitual) behavior).

[103] HARE, *supra* note 63, at 38, 129; *see also* PETER SINGER, PRACTICAL ETHICS 93 (2d ed. 1993) ("Even if we were to limit [our deliberations] to the more significant choices, there would be a danger that in many cases we would be calculating under less than ideal circumstances.").

[104] HARE, *supra* note 63, at 137; KAHNEMAN, *supra* note 27, at 105 ("We do not always think straight when we reason . . . .").

consequences (both good and bad) of general ways of acting in recurrent circumstances than to appreciate the expectable consequences of a single case." [105]

Habit also promotes efficiency. The alternative to reliance on habit is to conduct a case-specific balancing of the risks and benefits of each particular bodily movement and even, as Lewis's case illustrates, of each allocation of one's attention.[106] Balancing the risks and benefits associated with a course of action requires the actor, at the least, to identify the possible outcomes of the course of action, direct and indirect; to assign probabilities to those outcomes; and to assign values to those outcomes.[107] In short, the required calculation would "involve very complex thought processes."[108] As a consequence, a person who foreswore habitual behavior in favor of conducting such movement-specific evaluations would have trouble even getting dressed in the morning or washing his hands, as James said.[109] Driving a car would be out of the question.[110]

Accuracy and increased efficiency are not unrelated, moreover. They are complementary. Gains in efficiency often make themselves felt in improved accuracy. Take Lewis's case, of which we have stipulated (plausibly) that his inattention in the moment before the accident was the product of a habit. If Lewis, instead of relying on habit, had relied on conscious deliberation in deciding how to allocate his attention—if he had devoted some of his "limited stock of attention resources"[111] to deliberating consciously on the optimal allocation of his attention in each moment—he would not have had any attention resources left over for driving hazards. Experience in driving—the development of driving *habits*—generates

---

[105] BRATMAN, *supra* note 13, at 88.

[106] *See* HARE, *supra* note 63, at 36 (using car-driving to illustrate the impossibility of "meet[ing] each new situation entirely unprepared, and perform[ing] an 'existential' choice or cost-benefits analysis on the spot").

[107] *See* United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947) (articulating the Learned Hand formula for risk assessment).

[108] HARE, *supra* note 63, at 111.

[109] JAMES, *supra* note 22, at 5; *see also* BRATMAN, *supra* note 13, at 2–3 (explaining why, for the sake of efficiency, human beings "need ways for deliberation and rational reflection to influence action beyond the present"); SINGER, *supra* note 103, at 92–93 ("It is simply not practical to try to calculate the consequences, in advance, of every choice we make.").

[110] *See* HARE, *supra* note 63, at 36.

[111] CHABRIS & SIMONS, *supra* note 53, at 24.

increased safety in part by increasing the amount of attention the driver can devote to road.[112]

Because habits, like court rules, produce very substantial benefits in the form both of increased efficiency and of improved accuracy, human beings often are justified in cultivating habits in spite of the habits' inevitable over-inclusiveness.[113] Whatever the differences between habits and rules, then, habits are like rules in the only sense that matters to us: it is rational for an actor to cultivate and to retain habits even if habits occasionally prove overinclusive in specific cases—even if habits occasionally generate conduct whose risks outweigh its benefits.[114] And if it is rational for actors to cultivate habits, then it must be rational for them to act habitually too.[115] Bear in mind, what I'm claiming is not that the rationality of the pertinent habit—of the habit that generated the action—makes the actor's habitual action rational (or non-wrongful) from a time-slice perspective. When the risks associated with the actor's conduct outweigh the benefits, then the conduct still is wrongful, even if it was generated by a rational habit. This time-slice irrationality—this wrongfulness—exists in tension, though, with temporally extended rationality. This tension cannot be dissolved, either.[116] The conduct just is

---

[112] *See* Marianne R. Day, Andrew R. Thompson, Damian R. Poulter, Christopher B. Stride & Richard Rowe, *Why do Drivers Become Safer Over the First Three Months of Driving? A Longitudinal Qualitative Study*, 117 ACCIDENT ANALYSIS & PREVENTION 225, 225 (2018) ("General models of skill development propose that perceptual-motor performance becomes faster and more automatic with practice, making fewer demands on attentional resources.") (citation omitted).

[113] *See* Moore & Hurd, *supra* note 21, at 190 ("One does best . . . to develop heuristics, proxies, mottos, habits, and practices that allow one to make good decisions in circumstances that require hasty judgments, and that off-set, compensate for, or otherwise prevent the realization of one's own worst tendencies.").

[114] *See* BRATMAN, *supra* note 13, at 52 (discussing standards for evaluating the rationality of habits "in terms of their long-range consequences"); HARE, *supra* note 63, at 47–48 (discussing standards for evaluating the rationality of what Hare calls "prima facie principles," e.g., "the principle of always wearing our seat belts when driving"); *cf.* Hodgson, *supra* note 87, at 663 (observing that the differences between habits and rules, though "important," are not central to the critical question, namely, "Under what circumstances is it necessary or convenient for an agent to rely on habits or rules?").

[115] *See* Moore & Hurd, *supra* note 21, at 190 ("While the most extreme circumstances might justify one in second-guessing [heuristics, proxies, mottos, habits and practices], in most ordinary circumstances, one does best—that is, one acts rationally—only if one follows them as though they demanded blind allegiance.").

[116] *See* SCHAUER, *supra* note 86, at 98 (acknowledging the existence of a "chasm" between "rule-sensitive particularism," where decisions ultimately are evaluated on the basis of their time-slice rationality, and genuine "rule-based decision-making").

irrational from a time-slice perspective and is rational from the perspective of temporally extended agency.[117]

In a moment, we will turn to the question of how this tension is reflected in criminal law's twofold conception of liability. Before that, however, let's pause to consider a possible shortcoming in the rule–habit analogy.

D. HOW HABITS CONTROL CONDUCT

At first glance, the analogy between a judge's evidentiary rulings and an individual's habitual actions might seem misconceived. Judges who apply evidence rules are *required* to apply the evidence rules; they are *bound*. Evidence rules are promulgated by the state's highest court, usually, or by the state legislature.[118] Trial court judges are expected to apply those evidence rules, and judges whose rulings contravene the evidence rules will be subject to reversal by the appellate courts, or at least will violate strong norms of judicial behavior.[119] By contrast, a person who engages in habitual conduct cannot be said to be bound by the habit in the same way. No legal authority, for example, prohibits him or her from engaging in conduct that contravenes the habit. Nor does any social norm. So he or she remains free to depart from the habit—to rely instead on a case-specific balancing of the conduct's risks and benefits.

This is not a crazy objection. Nor could we have avoided the underlying question merely by choosing a different analogy. The underlying question is: How can the existence of a rule—whether in the form of a habit or a legal rule or whatever—impart rationality to conduct that *isn't* rational from a time-slice perspective?[120] The legal analogy was helpful because it strongly suggested that circumstances do exist where a general rule *can* impart

---

[117] *See* Mintoff, *supra* note 95, at 893 ("[I]t may be that the rule or disposition which maximally promotes an agent's welfare requires actions which do not maximally promote her welfare.").

[118] *See* Thomas A. Bishop, *Evidence Rulemaking: Balancing the Separation of Powers*, 43 CONN. L. REV. 265, 275 (2010) (explaining that the power to enact court rules usually "resides with state judiciaries, whether by constitutional or statutory grant, or as an attribute of the court's inherent authority to order its process.").

[119] *See* Montana v. Polak, 422 P.3d 112, 117 (Mont. 2018) ("A district court is bound by the Rules of Evidence and applicable statutes . . .").

[120] *Cf.* Alexander, *supra* note 99, at 54 (posing the question of how the existence of a rule can impart rationality to conduct that isn't rational from a time-slice perspective: "How can it ever be right not to do the right thing?"); Heidi M. Hurd, *Why You Should Be a Law-Abiding Anarchist (Except When You Shouldn't)*, 42 SAN DIEGO L. REV. 75, 76 (2005) (arguing that the existence of a rule cannot impart rationality to a decision that isn't time-slice rational: "the rationality of following any given rule resides in one's confidence that one is acting on the balance of reasons for action—including the good reasons for following the rule—and not at all in the fact that there is a rule . . . ").

rationality to otherwise irrational conduct. But now we have to reckon with the possibility that the rationality of evidentiary rulings depends on features of that setting that are not present where, say, habits are concerned.

As a first step, let's be clear about how the existence of an evidence rule does and does not affect the rationality of an individual evidentiary ruling. It is tempting to suppose that the existence of the evidence rule affects the time-slice rationality of individual evidence rulings by altering the balance of the individual ruling's downstream costs and benefits. After all, evidence rules *do* sometimes generate very substantial benefits in the form of increased efficiency and increased accuracy, as we have seen. It seems, at first glance anyway, as though it would be a simple matter to incorporate these rule-generated benefits into the time-slice cost-benefit balancing for individual evidence rulings. And it seems as though, after the incorporation of these rule-generated benefits into the time-slice cost-benefit balancing, every individual evidence ruling based on a rational evidence rule would itself be time-slice rational.

Alas, this doesn't work.[121] The difficulty is the same one we encountered in the lapse cases, namely, that the benefits generated by rules over time are not reflected in time-slice cost-benefit analyses for particular instances of rule-following.[122] Consider, first, the efficiency benefits that rules generate by relieving the decision-maker of the responsibility for conducting individualized cost-benefit analyses. In theory, each application of a rule generates efficiency benefits by relieving the decision-maker of the responsibility for conducting an individualized cost-benefit analysis in that case. And, in theory, incorporation of these benefits into the cost-benefit analysis for that case would tip the balance in favor of following the rule—and so would make each instance of rule-following rational. To incorporate these efficiency benefits into the cost-benefit analysis, however, the decision-maker would have to *conduct* a cost-benefit analysis. And if the decision-maker conducts an individualized cost-benefit analysis, then the decision-maker necessarily foregoes the efficiency benefits of *not* conducting an individualized cost-benefit analysis. In short, we cannot simultaneously suppose both (1) that the decision-maker harvests the efficiency benefits of foregoing case-specific deliberation in a particular case and (2) that the decision-maker conducts an individualized cost-benefits analysis in which these efficiency benefits are incorporated.

---

[121] *See* SCHAUER, *supra* note 86, at 97–98 (explaining why "rule-sensitive particularism" cannot capture the benefits associated with "rule-generating justifications" like "decision-maker disability" and efficiency).

[122] *See supra* text accompanying notes 76–79.

Much the same sort of difficulty arises when we try to capture, as part of the time-slice cost-benefits analysis for a particular case, the net accuracy improvements generated by rules. To incorporate these rule-generated benefits into the time-slice cost-benefits analysis, the hypothetical decision-maker would have to calculate the degree to which his or her individualized analysis in that case was likely to be undercut by, say, bias or prejudice or just the inherent difficulty of making complex judgments on the fly. In making *this* calculation, however, the decision-maker would be beset by the very shortcomings whose likely effects he or she was trying to calculate. Entrusting the distrusted decision-maker with this calculation seems, if not logically incoherent, at least "psychologically counterintuitive," as Fred Schauer has said.[123] "If we do not trust a decision-maker to determine *x*, then we can hardly trust that decision-maker to determine that this is a case in which the reasons for disabling that decision-maker from determining *x* do not apply or are outweighed."[124] For just this reason, Schauer argues, even "rule-sensitive particularism" cannot capture the benefits associated with "rule-generating justifications" like "decision-maker disability."[125]

Roughly the same thing is true where habits are concerned. Like the benefits generated by rules, the benefits generated by habits—increased efficiency, increased accuracy, etc.—would not make themselves felt in a hypothetical case-specific deliberation on the costs and benefits of *acting* on the habit, since they depend on non-deliberation. Granted, the existence of the habit might affect the actor's deliberations in one respect. A person who is deliberating about whether to deviate from a desirable habit in a particular instance—say, by smoking a cigarette, or by eating a donut for breakfast—would do well to consider the possibility that the departure will lead over time to diminution in the habit's force, as William James argued.[126] This investment in one's habits would be pointless, though, but for the *other* benefits of habit—the efficiencies generated by resolving behavioral dilemmas wholesale, for example, and the net improvement in decision-making accuracy over time. Since these other benefits depend on *foregoing*

---

[123] SCHAUER, *supra* note 86, at 98.

[124] *Id.*

[125] *Id.*

[126] JAMES, *supra* note 21, at 54–56.. James identifies two "great [practical] maxims" on the subject of habit. The first is that the actor should seize every opportunity to strengthen a desirable habit by performing the act he hopes to make habitual. *Id.* The second is that the actor should "[n]ever suffer an exception to occur till the new habit is securely rooted in your life. Each lapse is like the letting fall of a ball string which one is carefully winding up; a single slip undoes more than a great many turns will wind again." *Id.*

deliberation, they again could not make themselves felt in a hypothetical case-specific deliberation on whether to act in accordance with the habit.

If habits and rules do not affect the downstream costs and benefits that are determinative of conduct's time-slice rationality, then how can they make the conduct rational? The answer appears to be: By generating the conduct directly. When a rule or habit is rational from a temporally extended perspective—where the rule's or habit's costs in over-inclusiveness are outweighed by the rule's or habit's efficiencies—the rule or habit can impart rationality to behavior by controlling the behavior *directly*, without mediation by time-slice deliberative processes.[127]

Take our hypothetical evidentiary ruling, where the trial judge excluded the defendant's evidence of another suspect's character. The reason why we didn't concern ourselves with the time-slice rationality of the judge's ruling—with the question whether the evidence's probative value outweighed its potential prejudicial impact—is that the judge's ruling was controlled by the character rule, not by a time-slice balancing. The character rule, together with norms of judicial behavior requiring trial judges to apply their jurisdiction's duly promulgated evidence rules, took the issue out of the trial judge's hands, so to speak. Once we ascertained that the character rule controlled the judge's ruling, our analysis of the ruling's rationality naturally shifted from the ruling itself to the rule the judge was applying. The question became whether the rule was rational or cost-justified. What appeared to justify this shift in focus, though, was the *control* exerted by the character rule.

This, finally, brings into focus the concern that is the subject of this subsection: Do habits exert the required sort of control?

For starters, they do exert control of a kind. One of the features that distinguishes habitual behavior from other sorts of rule-following is the "self-actuating or autonomic quality" of habitual behavior.[128] Habits generate

---

[127] *Cf.* BRATMAN, *supra* note 13, at 16 (explaining that future plans are "conduct-*controlling* pro-attitudes": "My intention will not merely influence my conduct, it will control it"); Emily Sherwin, *The Rationality of Promising*, 27 CORNELL J. L. & PUB. POL'Y 533, 538 (2018) ("Traditional standards of practical rationality refer to the agent's current reasons for action at the time she acts. A number of writers, however, have proposed that practical rationally should be understood as extending over time, to allow for advance planning. David Gauthier, Edward McClennen, and Scott Shapiro all have proposed that, to realize the instrumental benefits of planning, agents must to some degree be bound by their prior intentions.").

[128] Hodgson, *supra* note 87, at 664; *see also* Simplex, Inc. v. Diversified Energy Sys., 847 F.2d 1290, 1293 (7th Cir. 1988) (identifying as one of the distinguishing features of habit that

behavior automatically and directly, without intervention of the conscious, deliberative mind—"without any consciously formed purpose or anticipation of results."[129] "Once the response is triggered in mind, people tend to act on it through ideomotor processes whereby the thought of a behavior brings about the corresponding physical response in a reflexive, automatic manner."[130] This is not to say that individuals can't ever override their habits in favor of case-specific deliberation. It's just to say that they often do not. As a result, habits often generate conduct more or less directly.

Would *this* sort of control, if exerted by a rational, cost-justified habit, suffice to impart rationality to habitual behavior? It would, I think. Philosopher Michael Bratman has addressed the related question whether, and when, advance planning can impart rationality to actions subsequently undertaken in accordance with the plans.[131] As it turns out, planning poses a riddle very like the riddle posed by habit. Advance planning does appear sometimes to bear on our judgments about the rationality of the actor's subsequent conduct.[132] Exactly how it affects the rationality of the conduct is somewhat mysterious, though.[133] On the one hand, it does not provide *reasons* for the actor later to engage in the planned conduct.[134] In other words, as with habit, the existence of the plans does not generally affect the balance of the conduct's risks and benefits in the moment of the conduct. At the same time, though, my advance plan "does not reach its ghostly hand over time"

---

it "is 'semi-automatic' in nature"); FED. R. EVID. 406, Advisory Committee's Note ("[T]he doing of the habitual acts may become semi-automatic.").

[129] JAMES, *supra* note 21, at 24 (quoting WILLIAM B. CARPENTER, PRINCIPLES OF MENTAL PHYSIOLOGY 344 (1874) ("[A]ny sequence of mental action which has been frequently repeated tends to perpetuate itself; so that we find ourselves automatically prompted to *think*, *feel*, or *do* what we have been before accustomed to think, feel, or do, under like circumstances, without any consciously formed *purpose* or anticipation of results."); *see also* MARCEL PROUST, SWANN'S WAY 117 (Christopher Prendergast, ed., Lydia Davis, trans., Penguin Books 2002) (1913) ("And from that moment on, I would not have to take another step, the ground would walk for me through that garden where for so long now my actions had ceased to be accompanied by any deliberate attention: Habit had taken me in its arms, and it carried me all the way to my bed like a little child.").

[130] Wood, Mazar & Neal, *supra* note 20, at 591; *see also* JAMES, *supra* note 22, at 6 (observing that "habit soon brings it about that each event calls up its appropriate successor without any alternative offering itself").

[131] BRATMAN, *supra* note 13, at 2–3.

[132] *Id.* at 23.

[133] *Id.*

[134] *Id.* at 26–27.

and control my future action as a present volition does.[135] Obviously, I can reconsider my plan.

To vastly oversimplify, Bratman's basic answer to this riddle is that advance planning influences our future conduct through the medium of "habits of nonreconsideration."[136] Though we *can* reconsider our plans when the moment of the planned conduct arrives, we usually do not.[137] Nor could we without sacrificing the considerable benefits, over a lifetime, of making advance plans and deliberating beforehand.[138] To preserve the benefits of advance planning and advance deliberation—to "enable us to avoid being merely time-slice agents"[139]—we develop habits of non-reconsideration, which preempt case-specific redeliberation in most, but not all, cases.[140] The control or influence exerted by advance plans via the medium of habits of non-reconsideration suffices, sometimes, to impart rationality to future actions.[141] Roughly, an action generated by a plan is rational if (1) the plan itself was rational and (2) the actor's habits of non-reconsideration also were rational, that is, were the kinds of habits we want "to encourage and develop in ourselves and others over the long run."[142]

The implications for us of Bratman's "planning" analysis are clear, in part because he more or less draws them himself. After explaining how an individual's advance plans can impart rationality to subsequent conduct that is not rational from a time-slice perspective, Bratman makes the same point in relation to "personal policies," which he says are a species of future-directed intention.[143] They are just intentions that "are general and concern potentially recurring circumstances in the agent's life."[144] Though Bratman

---

[135] *Id.* at 5.

[136] *Id.* at 60–70.

[137] *See id.* at 64 (identifying instances of "nonreflective (non)reconsideration" as "exceedingly common" and as central to agent rationality).

[138] *Id.* at 2.

[139] *Id.* at 34–35.

[140] *Id.* at 66–67 ("[G]iven our limits and the important role of plans in reliably extending the influence of present deliberation to future action we may expect that reasonable habits of (non)reconsideration will involve a tendency not to reconsider a prior plan except when faced with some problem for that plan.").

[141] *Id.* at 80.

[142] *Id.* at 70, 80; *see also* Mintoff, *supra* note 95, at 899–900 ("[I]f one rationally forms an intention to act later (perhaps if some condition is satisfied), and rationally does not reconsider that intention up to the time of action, then it is rational for one to act at that time (if that condition is satisfied).").

[143] BRATMAN, *supra* note 13, at 87.

[144] *Id.*

does not explicitly identify habits as a species of personal rule, his examples of personal rules—"[b]uckle up seat belts when driving in a car," for example—plainly encompass habitual behaviors.[145] Like advance plans, he says, personal rules impart rationality to conduct generated by the rules where (1) the personal rule is itself rational; and (2) the habits of non-reconsideration that mediated the rule's influence on the actor's later behavior also were rational.[146]

Bratman appears to assume that personal rules usually will be the products of conscious deliberation.[147] But the analysis would not change if the personal rules, or habits, instead were acquired without deliberation. A person's habit of fastening her seat belt, for example, is no less rational for the fact that she developed the habit as a child and never stopped to deliberate on its costs and benefits.[148] In criminal law, rationality is not about process. It's about risks and benefits.[149] This is true, of course, when we evaluate the rationality of a particular act from a time-slice perspective—when we decide whether the act is wrongful, in other words.[150] The same thing is or ought to be true when we evaluate the rationality of habit from a temporally extended perspective. Indeed, Bratman himself acknowledges as much.[151] Habits are rational, says Bratman, when they are justified by "the long-term consequences of their internalization by limited agents like us."[152]

E. CAN CRIMINAL LAW, LIKE TORT, IGNORE TEMPORALLY EXTENDED RATIONALITY?

In the end, we are faced with a kind of paradox. Whether an actor's habitual conduct counts as blameworthy sometimes seems to depend on which of two seemingly irreconcilable perspectives we adopt. When we evaluate the conduct's wrongfulness, and likewise when we apply a tort negligence standard to the conduct, we adopt a "time-slice" perspective on

---

145 *Id.*

146 *Id.* at 91.

147 *See id.* at 90.

148 *See* Moore & Hurd, *supra* note 21, at 186 ("From an early age, each of us has drilled into us a set of rules which our parents hope will become 'automatic' or routine. Don't play with matches; don't run along a poolside; look both ways before crossing a street . . . . As we become older, additional rules are added through social osmosis.").

149 *See* Dressler, *supra* note 15, at 957 ("To determine justifiability [in connection with recklessness], we conduct a criminal law version of the Learned Hand formula for measuring civil negligence . . . ").

150 *See id.*

151 *See* BRATMAN, *supra* note 13, at 52, 66.

152 *Id.* at 52.

the conduct, in Bratman's phrase.[153] From this time-slice perspective, the rationality of the actor's conduct depends exclusively on the downstream risks and utilities of the particular act.[154] But the actor's rationality also can be evaluated from a "temporally extended" perspective, again in Bratman's phrase.[155] When we take this broader perspective, the rationality of the conduct depends on the costs and benefits of the habit or habits that generated the conduct.[156]

How is the criminal law to respond to this seeming paradox? Tort law's response is instructive. Though the Restatement acknowledges the existence of the paradox—that wrongful conduct might be part of a larger pattern of behavior that is entirely reasonable—it does not adjust its negligence standard in response.[157] Rather, it concludes simply that the larger pattern defined by the actor's conduct over an extended period of time "is a reality that the jury is not in a position to consider."[158] The Restatement's answer to the paradox is not entirely satisfying even to tort scholars, who have characterized this approach as creating a "pocket of strict liability" in the law of negligence.[159] It creates a pocket of strict liability by, in effect, imposing liability on actors whose conduct is rational from a temporally extended perspective —by disregarding entirely an "important component of a normative account of practical rationality."[160]

Granted, tort scholars seem generally to take the view, as does the Restatement, that this "pocket of strict liability" ultimately is acceptable within the law of tort. Three good reasons support this view. First, it's not crazy to think that tort law should force individuals to internalize the costs associated with their development of particular habits, at least in cases where those habits generate conduct that is negligent from a time-slice perspective.[161] Second, as tort scholars have recognized, requiring factfinders

---

[153] *Id.* at 78-79.

[154] *See* RESTATEMENT (THIRD) OF TORTS, *supra* note 65, § 3 cmt. e at 30–31.

[155] BRATMAN, *supra* note 13, at 78–79.

[156] *Id.* at 52, 70.

[157] *See* RESTATEMENT (THIRD) OF TORTS, *supra* note 65, § 3 cmt. k at 36–37.

[158] *Id.* at 37.

[159] Mark F. Grady, *Res Ipsa Loquitur and Compliance Error*, 142 U. PA. L. REV. 887, 896–97 (1994); *see also* Kenneth S. Abraham, *Strict Liability in Negligence*, 61 DEPAUL L. REV. 271, 288–89 (2012); Fennell, *Accidents and Aggregates*, *supra* note 64, at 2392; Rabin, *supra* note 63, at 432–33.

[160] BRATMAN, *supra* note 13, at 51.

[161] *See* Fennell, *Accidents and Aggregates*, *supra* note 64, at 2391–92 (observing that the negligence standard reasonably can be viewed "as a safe harbor from liability that the law extends to actors under certain instrumentally defined circumstances, not an inalienable

both to identify the habits that generated the conduct and to evaluate their reasonableness would create very substantial administrative burdens.[162] Third, given the irreconcilability of the time-slice perspective and the temporally extended perspective, the only way to avoid the imposition of strict liability would be to treat the negligence standard as fundamentally twofold, rather than unitary.[163]

These justifications do not hold up where criminal law is concerned. For starters, strict criminal liability cannot be justified as a way of forcing individuals to "internalize" the costs associated with the adoption of habits. Unlike tort law, criminal law doesn't price; it punishes.[164] As nearly everyone agrees, punishment can't be justified in the absence of genuine moral blameworthiness.[165] And it is hard to claim that an actor's wrongful conduct demonstrates genuine moral blameworthiness where the actor was reasonable both in committing a particular sphere of his or her behavior to habit and in developing or retaining the very habits that he or she developed or retained.

In the criminal law setting, moreover, the fact that rationality over time cannot be collapsed into time-slice rationality, but rather poses a fundamentally distinct question, is no reason for disregarding it. Unlike tort law, criminal liability always has been thought to encompass two "conceptually distinct components"—wrongdoing and culpability, which *do not* collapse into one another.[166] The discovery that blameworthiness in cases like Lewis's hinges on a question that cannot be assimilated to our traditional conception of wrongdoing is, then, merely a reason for thinking that we have stumbled onto an aspect of the separate question of culpability. Indeed, given the long-standing and vexing uncertainty about the nature of culpability, we

---

human right to inflict all the harm on others that due care would not prevent (or even all the harm that one cannot personally help causing)").

[162] *See* Abraham, *supra* note 159, at 288–90 (discussing both "the cost of acquiring the information" relevant to this question and the dangers associated with the introduction of character and habit evidence at trial); Grady, *supra* note 64, at 402 ("For a court to lengthen the amount of time over which it defines the required rate of precaution essentially involves taking character evidence, which could be extremely self-serving to the defendant.").

[163] *Cf.* RESTATEMENT (THIRD) OF TORTS, *supra* note 65, § 3 cmt. k at 37 (seemingly acknowledging irreconcilability of the two perspectives by asserting that "the fallibility of average persons over a period of time is a reality the jury is not in a position to consider").

[164] John C. Coffee, Jr., *Does "Unlawful" Mean "Criminal"?: Reflections on the Disappearing Tort/Crime Distinction in American Law*, 71 B.U. L. REV. 193, 194 (1991) ("[T]ort law prices, while criminal law prohibits.").

[165] *See* THOMAS M. SCANLON, WHAT WE OWE TO EACH OTHER 266 (1998) ("Punishment is . . . an expression of 'legal blame.' Insofar as this is so, it will seem inappropriate, and hence to condemn, someone whose conduct is admitted to be blameless . . . .").

[166] Dillof, *supra* note 38, at 1029.

have reason to be encouraged by the discovery of a clearly defined question that (1) plainly bears on liability and (2) is rooted in the same fundamental principles as our test of wrongdoing but also (3) cannot be collapsed into the wrongdoing question.

We shouldn't get ahead of ourselves, of course. Though I will suggest later that culpability is fundamentally about temporally extended rationality, we are not there yet. It is possible that wrongdoing really is twofold, rather than unitary, and that we have stumbled onto a secondary aspect of wrongdoing, rather than a feature of culpability. It is also possible, of course, that culpability really is big, complex, and undefinable and we've merely succeeded in identifying a tiny aspect of this complex inquiry. So far, all I have suggested is that in cases like Lewis's, good habits seem to negate fault—that, in cases like Lewis's, the line between blameworthy and non-blameworthy conduct hinges on the difference between good habits and bad habits.

F. EVALUATING HABITS IN PRACTICE

If culpability is about the difference between good and bad habits, then how would the prosecutor go about proving culpability in, say, a case like Lewis's? In principle at least, the prosecutor ought to be able to prove culpability in either of two ways: (1) by proving that actor's habitual conduct was such that no hypothetical reasonable habit or set of habits could have generated it; or (2) by proving that the habit or set of habits that *actually* generated the conduct was unreasonable. In Lewis's case, as it turns out, the prosecutor appears to have used both avenues.

First, the prosecutor in effect proved that Lewis's conduct was such that no reasonable habit or set of habits could have generated it. Among the witnesses who testified at Lewis's trial was the driver of one of two logging trucks that were approaching from the other direction when the driver of the catering truck stopped to turn.[167] The second logging truck, as we know, struck the catering truck.[168] The driver of the first logging truck testified that, as he drove past the catering truck and then Lewis in the opposite direction, he was able to observe Lewis for approximately five seconds.[169] He testified that Lewis, "during that time, was looking down at his feet or lap and not at the road."[170] As the logging truck driver passed Lewis's truck, he saw Lewis

---

[167] State v. Lewis, 290 P.3d 288, 289 (Or. 2012).

[168] *Id.* at 290.

[169] *Id.* at 289.

[170] *Id.*

"suddenly look up, surprised and startled to see the catering truck stopped ahead."[171]

The logging truck driver's testimony that he was able to observe Lewis for five seconds as they passed seems somewhat implausible. After all, both trucks were traveling around the speed limit of 55 mph.[172] Indeed, though the trial court judge concluded that the logging truck driver was a credible witness, the judge also acknowledged that he "was unsure that [the driver's] estimate of 5 seconds was accurate."[173] Still, even if the other driver really had been able to observe Lewis only for two or three seconds, rather than five, his testimony still was strong evidence that Lewis at least "was inattentive for a protracted and significant period of time" just before the accident.[174]

So how does the length of Lewis's period of attention bear on his culpability if, as we have stipulated, his inattention was habitual rather than deliberate? It bears on his culpability by excluding the possibility that the inattention might have been generated by a reasonable habit of attention. We have acknowledged, of course, that even reasonable habits can prove overinclusive. Even reasonable habits, then, sometimes can generate conduct that is wrongful from a time-slice perspective. But it seems unlikely that any reasonable habit would cause a driver to spend two or three seconds looking down at his feet or lap while driving a tractor-trailer rig at 55 mph on a wet highway with an empty trailer. That is to say: The acknowledged benefits of habit—the increased efficiency and the improved accuracy—would not suffice to counterbalance the costs of generating this sort of protracted inattention.

The Oregon courts did not say this, of course. They just said, tracking the words of the Model-Penal-Code-derived Oregon statute, that Lewis's conduct represented a "gross deviation" from the standard of care that a reasonable person would observe under the circumstances.[175] This doesn't suggest that our analysis is wrong, though. The reasonable-person standard does not define an alternative standard with content of its own. It is an "empty

---

171 *Id.*

172 *Id.* Specifically, Lewis "had been traveling at 55 miles per hour (the speed limit for that section of the highway)," while the logging truck driver "was traveling about 45 miles per hour." *Id.* at 289-90.

173 *Id.* at 297.

174 *Id.*

175 *Id.* at 291 (quoting OR. REV. STAT. § 161.085(10)).

vessel," as one tort scholar has said.[176] It is a heuristic that serves multiple, not always well-defined objectives.[177] In the tort negligence context, for example, the reasonable person effectively is designed to aid the jury in the difficult exercise of balancing the downstream risks and benefits of the defendant's act or omission.[178] Within the criminal culpability context, it might well be that the reasonable person really is just a device for imagining the range of desirable, rational habits that might have been brought to bear under the circumstances.

Not all the evidence of Lewis's culpability was time-slice evidence. That is, not all the evidence was of his conduct in the instant before the crash or of the surrounding circumstances in that moment. The prosecutor also presented evidence of other conduct by Lewis from earlier that day.[179] Among the government's witnesses in *Lewis* were a husband and wife, Mr. and Mrs. Morrison, who testified that they had traveled behind Lewis's truck briefly when it was still about six or seven miles from the scene of the accident.[180] During this time, they observed several different incidents where Lewis's driving was, as the Oregon Supreme Court later said, "consistent with that of someone who was not paying attention to the road ahead of him."[181]

Was this evidence relevant?[182] The Oregon court thought so. It said that the Morrisons' testimony was admissible to prove Lewis's "ongoing state of

---

[176] Steven Hetcher, *Non-Utilitarian Negligence Norms and the Reasonable Person Standard*, 54 VAND. L. REV. 863, 864 (2001) ("The reasonable person standard is an empty vessel that jurors fill with community norms.").

[177] *See* Victoria Nourse, *After the Reasonable Man: Getting over the Subjectivity/Objectivity Question*, 11 NEW CRIM. L. REV. 33, 37 (2008) ("The reasonable person is a heuristic that, as we will see, serves multiple purposes.").

[178] RESTATEMENT (THIRD) OF TORTS, *supra* note 65, § 3 cmt. e at 30–31.

[179] *Lewis*, 290 P.3d at 290.

[180] *Id.*

[181] *Id.* at 298. The Morrisons saw Lewis make a last-minute un-signaled lane change to avoid being diverted into a small town business district, for example, and they saw him drive too closely behind a school bus. *Id.* at 290.

[182] The evidence would not, of course, have been admissible on a straight propensity theory. Or. R. Evid. 404(2), like Fed. R. Evid. 404(a)(1), bars the introduction of character evidence "for the purpose of proving that the person acted in conformity therewith on a particular occasion." This rule would have foreclosed use of the evidence for a propensity inference. That is, it would have foreclosed the inference that: (1) Lewis was driving irresponsibly a few minutes before the accident; (2) therefore, Lewis probably has a propensity for driving irresponsibly; (3) given this propensity, Lewis probably was driving irresponsibly in the moment of the crash too.

mind."[183] Specifically, the court said that the testimony was admissible to prove that Lewis's "inattention as he approached the catering truck was not just momentary or fleeting, but was instead the continuation of an ongoing distracted and inattentive mental state that characterized his driving for much or all of that morning."[184]

What the Oregon court meant by "state of mind" cannot be what lawyers usually mean by "state of mind."[185] A single period of "inattention," a single mental state of unawareness of the road in front of him, could not have spanned the entire period that encompassed the Morrisons' observations and the fatal accident. Nobody could drive six or seven miles without looking at the roadway. So Lewis's conduct over this period was characterized by several *different* periods of inattention. It does not seem likely, moreover, that all of these different periods of inattention could be attributable to a single "intention," say. It does not seem likely that in each of these moments of inattention, Lewis was consciously adhering to an overarching plan that required him to divert his attention at each of these moments individually.

So why does the court think that Lewis's earlier conduct tells us something about his culpability in the moment of the fatal accident? The best explanation is that Lewis's conduct in the earlier moment and his conduct in the moment before the accident were part of an overall pattern of attending too little or too infrequently to his driving. The fact that he had several close calls, and one fatal accident, within just a few minutes of driving provides strong, albeit probabilistic, evidence that the habit of attention that generated his conduct posed an unacceptable level of risk.

Notice, this theory of admissibility does not depend on an inference that Lewis's conduct in the instant before the crash was, in itself, the sort of conduct that is inconsistent with responsible driving. As we have seen, an occasional lapse in attention—even a lapse of the kind that gives rise to a fatal accident—is not necessarily inconsistent with being a responsible driver.[186] As these occasional lapses increase in frequency, however, so does the likelihood that one of the lapses will coincide temporally with a situation where even a brief lapse in attention will cause an accident. And as this likelihood increases, of course, the risk posed by the actor's habits of

---

[183] *Lewis*, 290 P.3d at 292.

[184] *Id.* at 298.

[185] *See* Gerard E. Lynch, *RICO: The Crime of Being a Criminal, Parts III & IV*, 87 COLUM. L. REV. 920, 932-33 (1987) (explaining that in "the common-law tradition" an "act or omission is conceived as taking place in an instant of time so precise that it can be associated with a particular mental state of intention, awareness of risk, or neglect of due care.").

[186] *Supra* text accompanying notes 63–75.

attention eventually becomes unacceptably great.[187] Because culpability can inhere in the overall pattern of the actor's conduct, as distinct from the nature of his conduct in a particular instant, testimony like the Morrisons' can be relevant to prove culpability even if it does not tell us anything about the actor's specific conduct in the instant of the crime.[188]

This, finally, is the second way in which the prosecutor could have proved Lewis's culpability. Not by proving that Lewis's conduct was such that no hypothetical reasonable habit could have generated it but, rather, by proving directly that the habit that *actually* generated the conduct was unreasonable. Mr. and Mrs. Morrisons' testimony about Lewis's earlier near misses, together with the fact of the fatal accident itself, suggested that Lewis's habits of attention were unreasonable, regardless of the exact nature of the lapse that caused the accident. [189]

II. TOWARDS A GENERAL THEORY OF CULPABILITY

So far, we have confined our attention to cases where the voluntary act or omission that provided the basis for the criminal prosecution was itself habitual rather than conscious. Most criminal cases aren't like this. In most criminal cases, the voluntary act that provides the basis for the prosecution will represent a conscious exercise of volition. In a typical prosecution for battery, for example, the act of striking or shoving the victim will be

---

[187] *See* Fennell, *Accidents and Aggregates*, *supra* note 64, at 2374–75 ("A single moment of inattention that produces an accident might be part of a larger pattern that represents as much care as any person could reasonably be expected to exercise. Alternatively, the accident-causing shortfall might be a representative draw from an urn of chronically unreasonable conduct.").

[188] *Cf.* Robert D. Cooter, *Punitive Damages, Social Norms, and Economic Analysis*, 60 LAW & CONTEMP. PROBS. 73, 88 (1997) ("Lapses in attention often cause unintended wrongs and under-appreciated risks. More vigilance by the actor reduces the frequency of lapses in attention.").

[189] One striking implication of this conclusion is that the same lapse might trigger liability in one case and not another. In the scenario where Lewis's liability hinges on his overall pattern of lapsing rather than on the character of the specific lapse that caused the accident, the lapse that caused the accident might, in itself, be no different from a nonculpable lapse. It might in itself be no different, that is, a lapse that occurs as part of a broader pattern of lapses that is nonculpable. Granted, if either of these two lapses were to trigger liability, they would both have to be wrongful—they both would have to pose unjustified risks from a time-slice perspective. Whether the lapses also were culpable, though, would depend on the overall pattern of lapsing—on the actor's habits—rather than on the character of the individual lapse. Professor Fennell seems troubled by this conclusion. *See* Fennell, *Accidents and Aggregates*, *supra* note 64, at 2383 ("If we tolerate liability for unavoidable lapses when they are mixed in with avoidable ones committed by the same person, why not otherwise?"). But she doesn't explore the issue in depth.

conscious rather than habitual.[190] Likewise, in a typical prosecution for murder, the act of shooting or stabbing the victim will be conscious rather than habitual.[191] And in a prosecution for drunk driving, the act of getting in one's car while intoxicated usually will be conscious rather than habitual. So do habits have any bearing on culpability in these more typical cases?

Another question left unanswered in Part I is whether the exculpating effect of desirable habits exhausts the content of the culpability requirement. Part I showed that criminal liability is inappropriate where the actor's conduct, though irrational from a time-slice perspective, partakes of the kind of rationality associated with desirable habits. It also suggested that what is lacking in these cases is culpability, rather than wrongdoing. But Part I did not show that the culpability is lacking *only* in cases where the actor's conduct is attributable to desirable habits. It seems possible that the culpability requirement, though it forecloses liability in these cases, forecloses liability in lots of other kinds of cases, too.

In this Part, I will try to suggest that: (1) though cases where the conduct itself is generated directly by habit are rare, cases where habit preempts some aspect of the actor's deliberative process are everywhere; and (2) except for the insanity defense, which operates on a different principle entirely, the existing law of culpability appears to be explained by the exculpating effect of desirable habits.

A. EXTENDING THE ARGUMENT: HABITS OF THINKING AND FEELING

Not all, or even most, habits generate behavior directly. Lots of habits—habits of "thinking," for example, and habits of "feeling"—influence the actor's behavior only indirectly, by influencing the deliberations that culminate in the actor's conscious exercise of volition.[192] In effect, habits of

---

[190] *E.g.*, Mooney v. State, 105 P.3d 149, 155 (Alaska Ct. App. 2005) (accepting defendant's explanation that his act of punching the victim in the head was an "instinctive, visceral, knee-jerk reaction," but declining to recognize defendant's conduct as "'involuntary' for legal purposes").

[191] *Cf.* Yaffe, *supra* note 56, at 177 (discussing hypothetical case where the defendant "has been trained by the military to spin around and fire immediately and without thinking on a threat behind him . . . " and posing the question: "Is he to be held guilty of a crime when, at the local firing range, he spins and fires on a person behind him who yells something threatening?").

[192] *See* Andrews, *supra* note 21, at 121 (defining habit as a "more or less fixed way of thinking, willing, or feeling acquired through previous repetition of a mental experience"); *see also* JAMES, *supra* note 21, at 24 ("[W]e find ourselves automatically prompted to *think*, *feel*, or *do* what we have been before accustomed to think, feel, or do, under like circumstances,

thinking and feeling cause the actor to deviate unconsciously from the kind of idealized, deliberative balancing of risks and benefits that is presupposed by the Hand formula. As psychologists Jennifer Robbennolt and Valerie Hans have said, with considerable understatement, "[t]he psychological research . . . makes clear that risk-utility balancing in strictest sense does not completely capture the nature of decision making about risk."[193] "Decision making about risk"—in other words, decision making of the kind that gives rise to criminal liability—"frequently happens quickly and unconsciously and is affected by a variety of heuristics," or "mental shortcuts."[194]

A habitual way of thinking might, for example, cause an actor to forego contemplation of a particular sort of risk.[195] Suppose, for example, that an actor decides to lend his car to a co-worker, despite indications that the co-worker is intoxicated or is a reckless driver. In this example, the actor's voluntary act of handing over his car keys obviously is conscious rather than habitual. At the same time, though, it would be easy to imagine a habitual way of thinking—a habit of trust, for example—affecting the actor's deliberations on the risk and thereby affecting, too, his ultimate decision to hand over the keys. Both psychologists and philosophers have acknowledged that trust often, and perhaps ideally, is unconscious: "we trust routinely, reflexively, and somewhat mindlessly across a broad range of social situations."[196] In our hypothetical, a habit of trust might have caused the actor unconsciously to forego deliberation on risks mediated by future misconduct of the co-worker. The actor's failure to deliberate on these risks might, in turn, have played a decisive role in the actor's ultimate decision to had over the car keys.

---

without any consciously formed *purpose*, or anticipation of results."); KAHNEMAN, *supra* note 27, at 24 ("System 1 continuously generates suggestions for System 2: impressions, intuitions, and feelings. If endorsed by System 2, impressions and intuitions turn into beliefs and impulses turn into voluntary action.").

193 ROBBENNOLT & HANS, *supra* note 18, at 41.

194 *Id.* at 41–42, 3.

195 *See* Heidi M. Hurd & Michael S. Moore, *Negligence in the Air*, 3 THEORETICAL INQ. L. 333, 386 (2002) (explaining how "heuristic or proxies" shape actors' assessments of risk in cases giving rise to tort or criminal liability).

196 Roderick Kramer, *Rethinking Trust*, HARV. BUS. REV. 69, 71 (2009); *see also* Annette Baier, *Trust and Antitrust*, 96 ETHICS 231, 235 (1986) (acknowledging that "there is such a thing as unconscious trust"); DORIS BROTHERS, FALLING BACKWARDS: AN EXPLORATION OF TRUST AND SELF-EXPERIENCE 3 (1995) ("Silent and invisible, trust rarely occupies the foreground of conscious awareness. We are no more likely to ask ourselves how trusting we are at any given moment than to inquire if gravity is still keeping the planets in orbit.").

Habits of thinking can influence an actor's deliberations in other ways, too. For example, they might influence the values assigned by the actor to possible outcomes. Most parents, for example, habitually would assign extraordinary weight to the interests of their children in calculating the value associated with possible outcomes.[197] Granted, this is not really a shortcut. It does, however, represent a habit-driven departure from the idealized Hand-formula risk calculus; the actor's children are not, after all, more valuable than any other person.[198] Habits of thinking might also cause an actor to ignore entirely some possible alternative courses of action in conducting his or her deliberations. For example, an avowed Kantian might automatically, and unconsciously, exclude from her deliberations any alternative course of action that put another person at risk purely for the actor's own benefit.[199]

Like habitual ways of behaving, habitual ways of thinking and feeling can be entirely rational even when they are overinclusive—even, that is, when they generate inputs that are irrational from a time-slice perspective. As "finite, limited beings" with "limited attention" and "limited cognitive power" too,[200] human beings cannot really afford to conduct complex risk-benefit calculations as a prelude to every action.[201] In most cases, they have little choice but to rely at least in part on automatically-generated "impressions and feelings"—on what psychologist Daniel Kahneman calls "System 1" thinking.[202] If reliance on System 1 is "the origin of much that we do wrong, . . . it also is the origin of most of what we do right," as Kahneman says.[203]

---

[197] *See* Agule, *supra* note 23, at 1013 (observing that "heightened sensitivity to [one's] children's interests . . . " is "largely constitutive of being a good parent").

[198] *See* MODEL PENAL CODE § 3.02 cmt. 3, at 14–15 (explaining that in the application of the choice-of-evils defense, "[t]he life of every individual must be taken . . . to be of equal value").

[199] *See* Richard W. Wright, *The Standards of Care in Negligence Law*, in PHILOSOPHICAL FOUNDATIONS OF TORT LAW 249, 256 (David Owen ed., 1995) ("[G]iven the Kantian requirement of treating others as ends rather than merely as means, it is impermissible to use someone as a mere means to your ends by exposing him (or his resources) to significant foreseeable unaccepted risks, regardless of how greatly the benefit to you might outweigh the risk to him.").

[200] Agule, *supra* note 23, at 1009.

[201] *See* HARE, *supra* note 63, at 36.

[202] KAHNEMAN, *supra* note 27, at 4 ("As we navigate our lives, we normally allow ourselves to be guided by impressions and feelings, and the confidence we have in our intuitive beliefs and preferences is usually justified."); *id.* at 20 (explaining that "System 1 operates automatically and quickly, with little or not effort and no voluntary sense of control.").

[203] *Id.* at 416; *cf.* Kenneth W. Simons, *When is Strict Criminal Liability Just?*, 87 J. CRIM. L. & CRIMINOLOGY 1075, 1099-100 (1997) ("The reasonably mistaken actor has, in effect,

Do habits of thinking and feeling bear on the actor's culpability? They do, though the relationship between the actor's habits and his or her culpability is more complex where habits of thinking and feeling are concerned than where action-generating habits are concerned. Where habits of thinking and feeling preempt some aspect of the actor's deliberations, the rationality of the operative habits would not itself absolve the actor of blame. In these cases, after all, the habits do not generate the conduct directly. They merely supply one of the inputs to the deliberative process. Even if the habit that generated a particular deliberative input was rational and desirable, then, we would still need figure out how the substitution of this habit-generated input affected the time-slice rationality of the actor's conduct. That is, we would have to recalculate the time-slice rationality of the conduct after substituting the habit-generated input for the usual Hand-formula risk inputs.

This sounds more complicated than it is. In the following three sections, I will try to show that this basic formula provides an intuitively plausible explanation for much of the existing law of culpability. It explains the duress defense and heat-of-passion defenses, for example. It also explains some aspects of proximate-cause doctrine.

B. NEGLIGENCE, RECKLESSNESS, AND THE HABIT OF TRUST

If we wanted to look for evidence that culpability is about habits of thought and feeling, where would we start? One promising alternative is reckless and criminally negligent homicide, where the factfinder's assigned responsibilities specifically include "mak[ing] the culpability judgment."[204] If culpability really were about habits, we would expect to encounter reckless and negligent homicide cases where the actor's habits of thought appeared to have informed the factfinder's culpability judgment. These habits would, presumably, be habits of the sort that affect the actor's assessment of risk, since recklessness and negligence are about the perception of risk.

I have already mentioned one habit that can influence an actor's perception of risk: the habit of trust. Risk sometimes is mediated by the future actions of other people.[205] When it is, naturally, the actor's perception of risk

---

adopted an epistemic 'system' analogous to the production system of the company. The epistemic system will predictably lead to some errors, but it is, by hypothesis, a reasonable system for which no better alternative exists.").

[204] MPC COMMENTARIES, *supra* note 10, at 238, 241.

[205] *See* 2 WAYNE R. LAFAVE, SUBSTANTIVE CRIMINAL LAW § 13.2(e), at 480–85 (discussing direct and accomplice liability for harms mediated by another person's misconduct); MPC COMMENTARIES, *supra* note 10, at 263–64 (discussing causal standards for cases where "volitional human intervention" mediates the causal connection between the actor's conduct and the harm).

often will depend on his or her judgments about the likely future conduct of other people. Judgments about others' future conduct sometimes are rooted in trust rather than evidence.

To illustrate: Suppose Dora is approached at work by a trusted co-worker, Abel, who asks if he can borrow her car the next day to take his daughter to the beach. Dora agrees. She tells Abel she will stop by his apartment that evening to drop off the keys. When Dora drops off the keys at Abel's apartment, Abel and several friends are celebrating a friend's birthday. Dora notices that Abel appears to be quite intoxicated. She hands him the keys anyway. An hour later, Abel strikes and fatally injures a pedestrian while making a beer run in Dora's car.

Whether Dora's conduct satisfies the wrongdoing component of reckless or criminally negligent homicide probably will depend on the specific factual details of the case—on what Dora knows about Abel, for example, and on what she observed at his apartment.[206] Still, it is plausible to suppose that under the circumstances known to Dora, the risks posed by giving Abel the keys outweighed the utilities. After all, none of the benefits Dora hoped to realize by her conduct depended on giving Abel the keys that evening rather than the next morning. Merely through the expedient of telling Abel an artful lie—"I just stopped by to tell you I won't be able to drop off the keys until tomorrow morning" — Dora could have eliminated most of the risk while still securing the hoped-for benefit of a day at the beach for Abel and his daughter.

If it is plausible to suppose that Dora's conduct was wrongful, though, it is also plausible to suppose that Dora's wrongful conduct was attributable to a rational habit of trusting her friends and acquaintances. Trust in others is "useful, valuable, important, or required" across a wide range of human relationships and endeavors, as philosopher Pamela Hieronymi has said.[207] The extraordinary social utility associated with trust does not, however, make itself felt in the time-slice balancing of risks and benefit. As Hieronymi has argued, the general "importance of trusting" in human life cannot provide a reason for trusting someone in a particular instance.[208]

---

[206] *See* MODEL PENAL CODE § 2.02(2)(c), (d) (AM. L. INST. 1985) (requiring the factfinder in recklessness and criminal negligence cases to assess the risk posed by the actor's conduct on the basis of "the circumstances known to him").

[207] Pamela Hieronymi, *The Reasons of Trust*, 86 AUSTRALASIAN J. PHIL 213, 213 (2008).

[208] *Id.* at 213–14; *see also* Carolyn McLeod, *Trust*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY, https://plato.stanford.edu/archives/fall2021/entries/trust [https://perma.cc/BPT4-RYW5] ("Considerations about how useful or valuable trust is do not bear on the truth of a trusting belief (i.e., a belief in someone's trustworthiness).").

> [T]o the extent that you must rely, for the justification or motivation of your trusting response, on reasons that appeal, not to the trustworthiness of the person in question, but rather to the importance of trusting [over the long term], to that extent you do not trust the person, and so, to that extent, your response was not (fully) trusting.[209]

So the rationality associated with trust is not time-slice rationality. It is temporally extended rationality. To reap the benefits associated with trust, then, we have to cultivate *habits* of trust. These habits of trust affect our deliberations without entering our conscious awareness, as psychologist Doris Brothers has said: "Trust rarely occupies the foreground of conscious awareness. We are no more likely to ask ourselves how trusting we are at any given moment than to inquire if gravity is still keeping the planets in orbit."[210] In Dora's case, then, a rational habit of trust might conceivably have preempted a critical aspect of her deliberations on risk. It might have caused her, rationally, not to consider any possible outcomes of her conduct that depended for their coming-to-fruition on future misconduct by Abel.

If this is right—if habits of trust really could defeat culpability—then we would expect to find evidence of this effect in the case law. Ideally, of course, this evidence would take the form of cases where courts say, explicitly, that culpability really is about rationality over time and that habits of trust sometimes make conduct rational. But courts do not say this, unfortunately. The case law does suggest, however, that something like this sort of reasoning is happening below the surface. Identifying habit's role in these cases takes some unraveling.

To begin with, courts do not appear ever to impose liability in cases like Dora's. Granted, there does not appear to be any reason in principle why conduct like Dora's should not give rise to direct, non-accomplice liability for reckless or negligent homicide. As Professor LaFave says: "[I]f A's own conduct in turning over the car to one known to be intoxicated is itself criminally negligent and if that conduct is found to be the legal cause of the death, then A is guilty of manslaughter on that basis."[211] In all the cases LaFave cites for this proposition, however, the owner was a passenger in the car when the fatal accident occurred.[212] The cases LaFave cites appear to be

---

[209] Hieronymi, *supra* note 207, at 214; *cf.* Baier, *supra* note 196, at 245 (remarking that "in general we cannot trust at will").

[210] BROTHERS, *supra* note 196, at 3.

[211] LAFAVE, *supra* note 205, at 485.

[212] *See id.* at 483 n.150 (citing State v. Hopkins, 265 P. 481, 481 (Wash. 1928) (upholding car owner's conviction as principal for manslaughter on negligent entrustment theory where car owner was a passenger in the car at the time of the accident)); *see also id.* at 485 n.155 (citing Freeman v. State, 362 S.W.2d 251, 253 (Tenn. 1962) (upholding car owner's conviction as principal for manslaughter on negligent entrustment theory where car owner was

representative, moreover. In cases where a vehicle owner is convicted of negligent or reckless homicide on the theory that he or she negligently or recklessly entrusted their vehicle to an intoxicated person, the vehicle's owner almost always is present in the vehicle as a passenger at the time of the fatal accident.[213]

So what does this feature of the case law signify? One interpretation of these cases is that when the car owner is not a passenger, he or she will not know enough about the background circumstances for her conduct to qualify as wrongful. After all, the risk posed by conduct is calculated on the basis of "the circumstances known to [the actor]."[214] The difficulty with this interpretation is that courts appear routinely to impose tort negligence liability in cases like Dora's.[215] Indeed, the Restatement (Third) of Torts includes an illustration very like Dora's case, where the actor "loans her car for the evening to her friend . . . who needs the car for social purposes."[216] Since the civil negligence standard and the wrongdoing component of criminal liability both are based on the same Hand-formula risk calculus, it seems unlikely that the trouble with criminal liability in Dora's case is about the wrongdoing component.

---

a passenger in the car at the time of the accident: "His permitting her to drive while intoxicated is the act that constituted criminal negligence on his part, especially when he knew the brakes on the car were defective.")); Hernandez v. State, 959 So. 2d 355, 356 (Fla. Dist. Ct. App. 2007) (upholding truck owner's conviction as a principal of "manslaughter by culpable negligence" where he was a passenger in truck at time of fatal accident).

[213] *See* J.A. Connelly, Annotation, *Who Other Than Actor is Liable for Manslaughter?*, 95 A.L.R.2d 175, § 10[a] 193–94 (1964). Though Connelly suggests that the "actual presence of the owner in the car is not a prerequisite to liability for manslaughter," *id.* at 193, the only case cited for this proposition is *Stacy v. State*, 306 S.W.2d 852, 853 (Ark. 1957), where the defendant truck owner was riding in another car "about two car lengths behind" the truck at the time of the fatal accident.

[214] MODEL PENAL CODE § 2.02(2)(c), (d) (AM. L. INST. 1985).

[215] *See* Frank J. Wozniak, Annotation, *Liability Based on Entrusting Automobile to One Who is Intoxicated or Known to be Excessive User of Intoxicants*, 91 A.L.R.5th 1, § 13[a] at 57-59 (2001) (collecting cases where courts imposed tort "liability on a friend for negligence in entrusting an automobile to a person whom the owner knew, or should have known, was intoxicated at the time of entrustment"); *cf.* DOBBS ET AL., *supra* note 79, at 367 ("[I]f the driver's dangerous movement of the vehicle causes harm, the defendant who entrusted the vehicle to a foreseeably dangerous driver cannot escape [tort] liability on the ground that the driver's negligence intervened to cause the harm, because the driver's negligence is the very danger he should have foreseen.").

[216] RESTATEMENT (THIRD) OF TORTS, *supra* note 65, at 218.

Professor Kadish has a different interpretation of the case law.[217] In effect, he argues that in cases like Dora's, where the car's owner is not a passenger, criminal liability is foreclosed by the causation requirement.[218] In particular, liability is foreclosed by the rule that the "free, deliberate, and informed intervention of a second person, who intends to exploit the situation created by the first, but is not acting in concert with him, . . . relieve[s] the first actor of criminal responsibility."[219] In Dora's case, he would argue, the drunk driver's free, deliberate, and informed decision to drive drunk cuts off Dora's liability.[220] Courts impose liability in cases where the car's owner also is a passenger only because they are surreptitiously treating the car's owner as an accomplice—as someone who intends to promote or facilitate the drunk driving.[221] Where the car owner is an accomplice, of course, the government need not prove a causal connection between the car owner's conduct and the result.[222] It need only prove causal connection between the principal's conduct and the result.[223]

At first blush, Kadish's account of these cases might seem not to have anything to do with culpability. After all, causation is an aspect of actus reus,

---

[217] Sanford H. Kadish, *Complicity, Cause and Blame: A Study in the Interpretation of Doctrine*, 73 CALIF. L. REV. 323, 401–02 (1985).

[218] *See id.* at 402 ("The difficulty is . . . finding the lending of the car to be the cause of death. The driver's actions in causing the fatal accident would have to be found sufficiently nonvoluntary to allow the causal inquiry to trace through his intervening actions in driving the car. This may be possible if the driver were drunk enough, but it may not be possible in many cases of driving under the influence.").

[219] H.L.A. HART & A.M. HONORÉ, CAUSATION IN THE LAW 326 (2d ed. 1985); *see also* Kadish, *supra* note 217, at 391 ("Actions are seen not as caused happenings, but as the product of the actor's self-determined choices, so that it is the actor who is the cause of what he does, not one who set the stage for his action.").

[220] *See* Kadish, *supra* note 217, at 401–02. Not all, or even most, courts accept the view that "free, deliberate, and informed" intervention by a second person invariably cuts off proximate cause. *See* MPC COMMENTARIES, *supra* note 10, § 2.03 cmt. 3 at 262 (declining to take a position on the question whether "volitional human intervention should be treated differently from other intervening causes"); United States v. Hatfield, 591 F.3d 945, 947–48 (7th Cir. 2010) (Posner, J.) (adopting a purely probabilistic view of proximate causation and making fun of phrases like "direct cause").

[221] MODEL PENAL CODE § 2.06(3)(a) (AM. L. INST. 1985) (defining scope of accomplice liability).

[222] *See* Commonwealth v. Smith, 391 A.2d 1009, 1011 (Pa. 1978) ("Once it has been determined that [the defendant] was an accomplice, proof that the principal caused the death satisfies the requirement of establishing the causal relationship of the accomplice.").

[223] *Id.*; *see also* Mark Kelman, *The Necessary Myth of Objective Causation Judgments in Liberal Political Theory*, 63 CHI. KENT L. REV. 579, 604 (1987) (remarking that the law governing accomplice liability "has traditionally (and wisely) disclaimed a need to find a causal nexus between conduct and consequence").

not mens rea. But this first impression is deceiving. It is true, of course, that cause-in-fact has nothing to do with culpability.[224] But the same is not true of proximate cause. Nowadays, just about everybody acknowledges that the proximate cause requirement is designed to distinguish (1) cases where the result represented the coming-to-fruition of the very risks that made the actor's conduct wrongful and culpable; and (2) cases where it did not.[225] When the proximate cause requirement blocks liability for a particular result, then, it is usually because the result flowed from a risk that was not wrongful, or was not culpable. In Dora's case, for example, if the proximate cause requirement blocks her liability for the results of Abel's drunk driving, it is because risks mediated by Abel's "voluntary, deliberate, and informed" misconduct were not among the risks that made her conduct wrongful or culpable.

This conclusion, finally, supports the view that culpability is about the rationality of an actor's habits. What distinguishes risks mediated by Abel's future misconduct from other risks from risks mediated only by the working of natural causes, say is that they're the very risks that a habitually trustful actor would overlook. And what distinguishes habits of trust from a million other habits of thought is that they are extraordinarily important to individual and community thriving and, so, often if not always are rational. The fact that tort law *doesn't* block liability in cases like Dora's only buttresses the obvious conclusion: When an actor's liability depends on genuine culpability, as it does in criminal law but not tort, an actor will not be liable if her conduct is reflective of temporally extended rational agency.[226]

---

[224] Justin D. Levinson & Kaiping Peng, *Different Torts for Different Cohorts: A Cultural Psychological Critique of Tort Law's Actual Cause and Foreseeability Inquiries*, 13 S. CAL. INTERDISC. L.J. 195, 214 n.103 (2004) ("While moral blameworthiness is relevant to a criminal law culpability determination, it is nonetheless a separate inquiry from the principle of actual cause.").

[225] *See* John C.P. Goldberg, Comment, *Misconduct, Misfortune, and Just Compensation: Weinstein on Torts*, 97 COLUM. L. REV. 2034, 2061 (1997) (explaining the "'risk rule' [approach to proximate cause], i.e., the rule that one should be held responsible only for harms flowing from the realization of the sort of risks that led society to regard the conduct as wrongful in the first place"); *see also* RESTATEMENT (THIRD) OF TORTS, *supra* note 65, § 29 (framing the proximate cause question as whether the plaintiff's injury "result[ed] from the risks that made the actor's conduct tortious").

[226] *See* Kadish, *supra* note 217, at 401 (discussing *Commonwealth v. Root*, 170 A.2d 310, 311–13 (Pa. 1961), where "[t]he court took note of several tort cases that had expanded proximate cause to include the foreseeable actions of others, but regarded it as inappropriate to apply them to criminal cases").

C. PROVOCATION AND THE HABIT OF PRIDE

A theory of culpability must, of course, explain the standard "excuse" defenses, which are thought to negate culpability.[227] Part of the reason why excuse defenses usually are said to negate culpability is that they obviously do not negate the other fundamental component of criminal liability, namely, wrongdoing.[228] Negating the wrongdoing component is the function of justification defenses, which defeat liability in just those cases where the risk posed by the conduct doesn't outweigh the utilities.[229] "Excuse" covers everything that is not a justification.[230] So excuses defeat liability where the actor's conduct (1) isn't time-slice rational, and therefore isn't justified; but at the same time (2) isn't deserving of condemnation either, or at least (like killings committed in the heat-of-passion) is less deserving of condemnation.

One reason why courts and scholars have struggled to explain the excuse defenses is that the defenses lumped together in this catchall category really operate on at least two fundamentally different principles, as A.P. Simester and others have observed.[231] Some excuse defenses—insanity, for example—operate on the principle that only morally responsible agents are appropriate objects of punishment.[232] Other excuse defenses, though, are triggered by circumstances that "have no tendency to make us see the agent as other than a morally responsible agent."[233] The circumstances that trigger

---

[227] BRINK, *supra* note 61, at 3 ("Justifications deny that the conduct in question is in fact wrong. . . . By contrast, excuses concede wrongdoing but deny that the agent is culpable or responsible for the wrong.").

[228] *See* GEORGE P. FLETCHER, RETHINKING CRIMINAL LAW 759 (2000) (explaining that justification defenses "challenge whether the act is wrongful," whereas excuse defenses address the question "whether the actor is accountable for a concededly wrongful act"); A.P. SIMESTER, FUNDAMENTALS OF CRIMINAL LAW 401 (2021) ("That division [between justification and excuse] is helpful, in as much as excuses directly engage findings of culpability whereas justifications are primarily concerned with wrongdoing.").

[229] *See, e.g.*, MODEL PENAL CODE § 3.02 (AM. L. INST. 1985) (defining choice-of-evils defense).

[230] *See* SIMESTER, *supra* note 228, at 401 (acknowledging that "the category of excuses [is] residual").

[231] *Id.* at 247 ("[T]here is a difference between defences that deny moral responsibility and defences that assert responsibility but deny culpability."); *see also, e.g.*, Peter F. Strawson, *Freedom and Resentment*, 48 PROCEEDINGS OF THE BRITISH ACADEMY 187, 192–94, 201–02 (1962) (identifying two different kinds of circumstances that inhibit the "reactive attitudes": (1) circumstances like profound mental illness that "invite us to suspend our ordinary reactive attitudes towards the agent" entirely; and (2) circumstances like duress that "inhibit indignation without in any way inhibiting the sort of demand on the agent of which indignation can be an expression").

[232] Strawson, *supra* note 231, at 193–94.

[233] *Id.* at 202.

the duress defense, for example, far from suggesting the actor is not responsive to reason, suggest rather that the actor's conduct is, in some slightly unusual sense, "reasonable."[234] Likewise, the "sudden provocation" that triggers the common law heat-of-passion defense is thought somehow to make the actor's emotional response, if not the killing itself, "reasonable."[235]

Even after we have recognized that excuses in this second category, like duress and heat of passion, operate on a different principle than does, say, the insanity defense, this second category remains problematic. The trouble is that it is unclear in exactly what sense the conduct excused by these defenses might be "reasonable." The criminal law generally conceives of reasonableness as time-slice rationality; it conceives of reasonableness exclusively as a feature of specific acts, rather than of extended patterns of behavior. Against this background, scholars understandably have endeavored to show that these defenses must partake somehow of time-slice rationality. They have argued, for example, that duress really is a justification defense, not an excuse, and that the heat-of-passion defense is a "partial justification" defense.[236]

These efforts have not fared well. As Professor Dressler has observed, neither the caselaw nor our intuition supports the claim that duress should be available as a defense only when the actor's conduct represents the lesser evil.[237] And the claim that heat of passion is a partial justification defense seems even less plausible.[238] For starters, the idea of "partial justification" is

---

[234] *See* MODEL PENAL CODE § 2.09 (AM. L. INST. 1985) (requiring, as an element of duress defense, that the threat of force be one "that a person of reasonable firmness in [the actor's] situation would have been unable to resist"); United States v. Lopez, 913 F.3d 807, 815 (9th Cir. 2019) ("The formula is addressed to the impact of a threat of force upon a reasonable person: The fear must be 'well-grounded.' There must be no 'reasonable' opportunity to escape.") (citing MODEL PENAL CODE § 2.09(i) (AM. L. INST. 1985).

[235] *See* People v. Casassa, 404 N.E.2d 1310, 1315–16 (N.Y. 1980) (discussing objective, reasonable-person component of the Model Penal Code's "extreme mental or emotional disturbance" defense to murder); Maher v. People, 10 Mich. 212, 220 (1862) (framing the objective question posed by the common law heat-of-passion defense as whether the provocation was such that it "*might* render ordinary men, of fair average disposition, liable to act rashly or without due deliberation or reflection").

[236] *See* Peter Westen & James Mangiafico, *The Criminal Defense of Duress: A Justification, Not an Excuse—and Why It Matters*, 6 BUFF. CRIM. L. REV. 833, 836 (2002); Mitchell N. Berman & Ian P. Farrell, *Provocation Manslaughter as Partial Justification and Partial Excuse*, 52 WM. & MARY L. REV. 1027, 1066, 1078-79 (2011).

[237] DRESSLER, *supra* note 62, at 285–86.

[238] *See* Eric A. Johnson, *When Provocation is No Excuse: Making Gun Owners Bear the Risks of Carrying in Public*, 69 BUFF. L. REV. 943, 974–75 (2021) (exploring shortcomings of the partial-justification account of provocation manslaughter).

itself problematic, much as the idea of "slight negligence" is problematic: conduct either satisfies the time-slice Hand-formula balancing of risks and benefits or does not; there is no in-between."[239] More importantly, though, in heat-of-passion cases the provoked actor's retributive aims don't actually provide any partial justification, or any "partially warranting reason," for killing another person in anger.[240] In modern societies, as Jeremy Horder observes, "it is the state that claims an all-embracing authority to act on . . . moral reasons relating to the justification for the deliberate infliction of considered punishment and retribution."[241]

If crimes committed under duress or in the heat of passion do not have any claim to time-slice rationality, though, in what sense can the actor's conduct or reaction be "reasonable"? The answer, as I will argue in this subsection and the next, is that defenses like these really are about temporally extended rationality.

Let's begin with the heat-of-passion defense. Both the heat-of-passion defense and its Model Penal Code counterpart, the "extreme emotional or mental disturbance" defense, have two components. Both defenses have a subjective component, which requires the actor to show that he or she killed the victim in a state of extreme emotional disturbance.[242] But both defenses also have an objective component. The heat-of-passion defense's objective component requires the actor to show (1) that the actor's extreme emotional state was engendered by a provocative event and (2) that the provocative event was of a kind that would have engendered an extreme emotional state in a "reasonable person," too.[243] The more-flexible Model Penal Code defense does not require the actor to prove the occurrence of an external provocative event.[244] It does, however, require the actor to show that his or

---

[239] Gary T. Schwartz, *Tort Law and the Economy in Nineteenth-Century America: A Reinterpretation*, 90 YALE L.J. 1717, 1743 n.203 (1981) ("If 'negligence' is itself given a rigorous economic definition, then 'slight negligence' seems either economically meaningless or economically wrong.").

[240] Johnson, *supra* note 238, at 975.

[241] JEREMY HORDER, PROVOCATION AND RESPONSIBILITY 176 (1992).

[242] *See* LAFAVE, *supra* note 205, at 672 (identifying subjective "emotional disturbance" as an element of common law defense); MODEL PENAL CODE § 210.3(1)(b) (AM. L. INST. 1985) (identifying "extreme mental or emotional disturbance" as a component of the Model Penal Code's counterpart to the traditional common law heat-of-passion defense).

[243] *See* LAFAVE, *supra* note 205, § 15.2(a) at 673.

[244] *See* MODEL PENAL CODE § 210.3 cmt. 5 at 56–56.

her extreme emotional state has an objectively reasonable explanation, either external or internal.[245]

Most of the controversy generated by these defenses, both among courts and among scholars, has centered on the objective component and, in particular, on the question what sort of "reasonableness" can be ascribed to the perpetrator of a violent, intentional, unjustified homicide. As we have already seen, some scholars have argued, implausibly, that what the heat-of-passion defense requires by way of reasonableness is a limited, "partial" sort of time-slice rationality.[246] Others have argued that the reasonableness of the actor's response depends on what the response tells us about the actor's moral character.[247] Finally, still others have argued that efforts to distinguish reasonable from unreasonable provocations should be abandoned as incompatible with the requirements of a "liberal state."[248]

There is a better answer. As we have already seen, an actor's rationality can take either of two fundamentally different forms. It can take the form of time-slice rationality, as it does when the conduct's risks are outweighed by its utility. But it also can take the form of temporally extended rationality, as it does when the actor's conduct is the product of habits whose benefits outweigh their costs. The "reasonableness" that is ascribed to the actor's emotional reaction in heat-of-passion homicide is temporally extended rationality, not time-slice rationality. In short, subjective emotional states of the kind that trigger the defense generally are the products of habitual moral reaction patterns. And the rationality of these habitual moral reaction patterns can be evaluated like the rationality of other habits of thought or feeling, namely, according to whether their benefits over time outweigh their costs.

As a first step, it is important to be clear about what generates the actor's subjective extreme emotional response. Courts and scholars sometimes speak as if the actor's emotional reaction, or even the actor's conduct, could be ascribed directly to the external provocation. They speak of the actor's conduct being "attributable to the distorting effect of surrounding

---

[245] *Id.*; *see also* People v. Casassa, 404 N.E.2d 1310, 1315–16 (N.Y. 1980) (discussing objective components of the Model Penal Code defense).

[246] *See* Berman & Farrell, *supra* note 236, at 1090 (defending the view that the provocation defense is a "partial justification" defense and explaining that the "degree of wrongfulness" of conduct – the degree of justification, in other words – depends on both "the harm that an actor intends or anticipates" and the countervailing "reasons on which conduct is based").

[247] *See* Johnson, *supra* note 238, at 976–80.

[248] *See* Stephen P. Garvey, *Passion's Puzzle*, 90 IOWA L. REV. 1677, 1716 (2005).

circumstances"[249] or to "external interference of certain kinds."[250] But this is, of course, just shorthand. Except in the rare case where the movements of the actor's body are attributable to physical manipulation by an outside force, the influence of external events is mediated by the actor's brain.[251] In the case of heat-of-passion homicide, for example, the influence of the provocative event is mediated by the actor's emotional reaction patterns. At least in the usual case, the actor does not decide consciously what to feel in response to, say, being punched in the face, or to catching her spouse in adultery.[252] Rather, the actor's emotions are generated by habitual reaction patterns—by habits of feeling, in James's formulation.[253]

It is possible to evaluate the rationality of habitual reaction patterns just as we evaluate the rationality of other habits, namely, according to whether their risks exceed their benefits. Habitual reaction patterns do, of course, produce real benefits. "[T]he formation in ourselves of relatively simple reaction-patterns" is, as Hare observed, "an indispensable help in coping with the world."[254] If not for such reaction patterns, "we should have to meet each new situation entirely unprepared, and perform an 'existential choice' or a cost-benefit analysis on the spot."[255] The efficiencies associated with these simple reaction patterns often are sufficient, moreover, to outweigh their occasional costs in overinclusiveness—to outweigh the risk that the reaction pattern will be triggered in a situation where it is unhelpful.

Even habitual reaction patterns that cause extreme anger sometimes are reasonable in this way. Suppose, for example, that the actor kills another person in anger after being punched in the face.[256] Probably the actor's extreme anger in this case has less to do with the injury inflicted than with

---

[249] FLETCHER, *supra* note 228, at 806.

[250] BRINK, *supra* note 61, at 4.

[251] *See* FLETCHER, *supra* note 228, at 811 ("Excuses apply on behalf of morally involuntary responses to danger; they acknowledge that when individuals merely react rather than choose do wrong, they cannot fairly be held accountable.").

[252] *See* JOSEPH E. LEDOUX, THE EMOTIONAL BRAIN: THE MYSTERIOUS UNDERPINNINGS OF EMOTIONAL LIFE 19 (1996) (observing that "conscious control over emotions is weak").

[253] JAMES, *supra* note 21, at 48; *see also* KAHNEMAN, *supra* note 27, at 105 (identifying the generation of "impressions, feelings, and inclinations" as one of the roles of "System 1" cognition, which "operates automatically and quickly, with little or no effort and no sense of voluntary control").

[254] HARE, *supra* note 63, at 36.

[255] *Id.*

[256] *See, e.g.*, Stewart v. State, 78 Ala. 436, 440 (1885) ("There can be no doubt of the fact, that a blow in the face with the hand, intentionally inflicted by one person on another, may constitute a provocation adequate to reduce a homicide from murder to manslaughter, even when such killing is perpetrated with a deadly weapon.").

the affront to the actor's pride or dignity.[257] (The actor would not react with the same sort of anger, probably, if the blow were inflicted accidentally, or by a child.) It is common to think of pride as a vice. And of course an undue concern with how one is perceived by others—whether in the form of, say, a too-punctilious sense of personal honor, or in the form of a crippling self-consciousness—*can* be vicious.[258] As Simon Blackburn argues in his recent book *Mirror, Mirror*, however, habitual reaction patterns of the kind we characterize as "pride" or "self-respect" also "may instead be something that is absolutely essential to protect, a prime motivator to good behavior . . . a guarantor of our integrity, a lifeline to which to cling."[259] Blackburn asks: "Might not an unself-conscious person . . . unself-consciously run away on the battlefield, or unblushingly pocket a dropped purse?"[260]

Of course, some reaction patterns that generate homicidal anger are not reasonable in this way. Take the case of Stephen Carr, who killed two lesbian hikers on the Appalachian Trail after becoming enraged by the "'show' put on by the women."[261] However we identify the emotional reaction pattern at work in Carr's case—as a pattern of reacting with anger to homosexual behavior, say, or as a general hostility toward women who are not interested in him sexually[262]—it is hard to identify any benefits that might be generated by the reaction pattern. So the risks posed by Carr's reaction pattern, including the risk of homicide, outweighed the benefits.

This is not what the court said, of course, in rejecting Carr's claim that he was entitled to a jury instruction on the heat-of-passion defense. The court did not go to the trouble of trying to identify the habitual reaction pattern that had generated Carr's anger, much less consider explicitly whether the risks posed by this reaction pattern were offset by its benefits. The court just said, as courts do, that the provocation was not of a kind that would have provoked

---

[257] *See* HORDER, *supra* note 241, at 30 ("All the important categories of sufficient provocation are explicable in terms of the importance attached [historically] to the concept of natural honour.").

[258] *See* SIMON BLACKBURN, MIRROR, MIRROR: THE USES AND ABUSES OF SELF LOVE 10 (2014).

[259] *Id.*

[260] *Id.* at 26; *see also* GEORGE ELIOT, MIDDLEMARCH (1872) *in* WORKS OF GEORGE ELIOT, vol. VI, p. 86 (Century Co. 1910) ("Pride helps us; and pride is not a bad thing when it only urges us to hide our own hurts—not to hurt others.").

[261] Commonwealth v. Carr, 580 A.2d 1362, 1363–64 (Pa. Super. Ct. 1990).

[262] *Id.* at 1363–64 ("In support of this defense and to show the existence of passion, Carr offered to show a history of constant rejection by women, including his mother who may have been involved in a lesbian relationship, sexual abuse while in prison in Florida, inability to hold a job, and retreat to the mountains to avoid further rejection.").

a reasonable person: "The sight of naked women engaged in lesbian lovemaking . . . is not an event which is sufficient to cause a reasonable person to become so impassioned as to be incapable of cool reflection."[263]

Though courts do not actually talk about desirable and undesirable habits in these cases, the courts' decisions are readily explicable in these terms. Even if the courts' analyses were habit-focused, the courts in most cases would not receive evidence of the defendant's broader reaction habits; they would just hear evidence about his reaction in the charged case. Evaluation of the actor's habits necessarily would be hypothetical, then. The question would just be whether *any* reasonable, justified habit could have generated an emotional reaction of the kind that led to the killing. This question, of course, is really just a minor elaboration of the question the courts actually consider: namely, whether a "reasonable person" would have been provoked.[264] It just defines the reasonable person as a person of reasonable, justified habits.

This is not to say, however, that courts wouldn't consider evidence of the defendant's *actual* reaction habits in the right case. If it is possible for the factfinder to reconstruct after the fact the reaction patterns that generated the anger, then the factfinder's evaluation of those reaction patterns ought to be decisive. Consider *Taylor v. State*, where murder-defendant Charles Taylor claimed that he had killed his girlfriend's secret lover in the heat of passion.[265] At Taylor's trial, the government introduced evidence of three previous incidents where Taylor had threatened homicidal violence in response to very minor incidents of sexual jealousy.[266] On appeal, the Wyoming Supreme Court held that these prior incidents were relevant to show that Taylor "engag[ed] in threatened violence even when not under extreme provocation."[267] The reaction pattern that generated Taylor's homicidal response to the provocation by his girlfriend's secret lover was not reasonable, in other words.

D. DURESS AND THE HABIT OF LOYALTY

Now consider duress. The duress defense varies slightly from jurisdiction to jurisdiction, but it basically requires the actor to show that another person threatened credibly to inflict death or serious bodily harm on

---

[263] *Id.* at 1364.

[264] *See* DRESSLER, *supra* note 62, at 504–05.

[265] Taylor v. State, 203 P.3d 408, 411, 413 (Wyo. 2009).

[266] *Id.* at 411.

[267] *Id.* at 413.

the actor or a third person if the actor did not commit the charged offense.[268] Both in its common law form and in its more flexible Model Penal Code form, the defense is shot through with demands for reasonableness.[269] Under the common law, for example, the actor must be reasonable both in crediting the threat and in concluding that he or she lacks an avenue of escape.[270] Under the Model Penal Code, the threat must be one that a "person of reasonable firmness" would have been unable to resist.[271]

What duress requires by way of reasonableness is not time-slice reasonableness. Even when an actor faces a credible threat of death or serious bodily injury, the actor's compliance with the threatener's demands rarely will "achieve the greater good" or "promote the general welfare," as the Ninth Circuit acknowledged in *United States v. Contento-Pachon*.[272] Take *Contento-Pachon* itself, for example, where the defendant, a native of Colombia, claimed that he had smuggled drugs into the United States only because drug traffickers had threatened to kill him and his family if he did not smuggle the drugs.[273] Though the court held that Contento-Pachon should have been permitted to raise the duress defense, it also said "he did not act to promote the general welfare."[274] This wasn't crazy, either. From a time-slice perspective, it is at least arguable that Contento-Pachon ought to have defied the drug traffickers. In the long run, after all, giving into the demands of drug cartels only strengthens their grip on the community.

So the duress defense does not require time-slice rationality—does not require the actor to show that the downstream utility of the charged criminal act outweighed the act's risks. From the fact that the duress defense does not require time-slice rationality, courts and scholars usually jump directly to the conclusion that the defense does not require any sort of rationality at all.[275] On this view, the duress defense merely is a "limited concession to human weakness."[276] The question it poses is not whether the actor's conduct is

---

[268] *See* DRESSLER, *supra* note 62, at 283–85.

[269] United States v. Lopez, 913 F.3d 807, 815 (9th Cir. 2019) ("The formula is addressed to the impact of a threat of force upon a reasonable person: The fear must be 'well-grounded.' There must be no 'reasonable' opportunity to escape.").

[270] *Id.*; *see also* DRESSLER, *supra* note 62, at 283–85.

[271] MODEL PENAL CODE § 2.09(AM. L. INST. 1985).

[272] United States v. Contento-Pachon, 723 F.2d 691, 695 (9th Cir. 1984).

[273] *Id.* at 693.

[274] *Id.* at 695.

[275] *See* DRESSLER, *supra* note 62, at 283–85 (urging that the question in duress cases isn't whether the harm perpetrated by the defendant represents a "lesser harm" than the harm threatened against him or her, but rather whether the defendant "should be blamed").

[276] Bergelson, *supra* note 9, at 317.

rational but, rather, whether the factfinder is capable of empathizing or identifying with the actor—whether, in the words of the Model Penal Code commentary, the factfinder is "not prepared to affirm" that they would act differently than the actor "if their turn to face the [same] problem should arise."[277] On this view, the only thing that distinguishes Mr. Contento-Pachon from, say, an actor who commits a murder to avoid a threatened punch on the nose is that one elicits a sense of identification in the factfinder and the other does not.[278]

Courts and scholars are wrong, though, to suppose that human rationality can only take the form of time-slice rationality. As we have seen, human rationality also can take the form of "temporally extended rational agency."[279] Just this sort of rationality, moreover, arguably characterizes actors who, like Contento-Pachon, commit crimes in response to serious threats against themselves or their family members. Most actors habitually assign greater weight to their own and their family members' interests than to the interests of strangers and of society as a whole. This habit of loyalty to oneself and one's family members is rational over the long term, moreover. As Hare has said: "If mothers had the propensity to care equally for all the children in the world, it is unlikely that children would be as well provided for even as they are."[280] Our "best bet," then, is to cultivate in ourselves a strong habitual preference for ourselves and our family members, even if, as in *Contento-Pachon*, this preference occasionally generates conduct that is not time-slice rational.[281]

A habit of loyalty might affect an actor's deliberations in at least a couple of ways. First, it might affect the *values* the actor assigns to various possible outcomes of the actor's conduct. It might, in short, cause the actor to assign greater weight to adverse outcomes that affect the actor and his or her family members than to outcomes that adversely affect strangers or society as a whole. Philosopher Craig Agule has argued that, indeed, the duress defense might be explained by the rationality of what he calls

---

[277] MPC COMMENTARIES, *supra* note 10, at 374–75.

[278] *Cf.* Westen & Mangiafico, *supra* note 236, at 909 (challenging the view that the factfinder's ability to identify with the actor ought to be determinative: "Yet it is not hypocritical for the judges to condemn actors for doing what the judges admit they, too, would do, if the judges candidly acknowledge that if they, too, did such a thing, they, too, would deserve to be condemned for it.").

[279] BRATMAN, *supra* note 13, at 78.

[280] HARE, *supra* note 63, at 137.

[281] *See id.*

"normative contour[ing]."[282] Normative contouring, as Agule describes it, occurs when an actor is "comparatively more sensitive to some reasons and comparatively less sensitive to others."[283] Normative contouring makes sense over the long term because, given human beings' limited cognitive resources, "an agent who is especially sensitive to regular, tractable, and important matters will get those matters right more often"—more often, that is, "than a uniformly sensitive agent will."[284] In cases like *Contento-Pachon*, then, the actor's conduct, though not time-slice rational, may reflect nothing more than "the proper functioning of a good normative psychology."[285]

Alternatively, a habit of loyalty might affect the actor's deliberations by removing some alternatives—some possible courses of action—from the table entirely. A habit might take the form, for example, of a perceived "duty" to protect one's family members even in the face of overwhelming danger to non-family members.[286] This sort of duty was implicated in philosopher Frank Chapman Sharp's original version of the trolley problem.[287] In Sharp's problem, a man is walking home across a railroad track when he notices that a switch has carelessly been left open.[288] The only way for him to save the passengers on a fast-approaching train is to close the switch.[289] In that moment, though, the man realizes that his own child is playing on the tracks and that by closing the switch he would send the train hurtling toward her.[290]

This is not, strictly speaking, a duress case, or even a criminal case, since an omission to close the switch will not expose the actor to criminal

---

[282] Agule, *supra* note 23, at 1010.

[283] *Id.* at 1008.

[284] *Id.* at 1010.

[285] *Id.* at 1021. Notably, Agule supposes—wrongly, in my view—that to the degree that action under duress represents the proper functioning of good normative psychology, the duress defense can't be regarded as one of the "ordinary excuse[s]." *Id.* On the contrary, I would argue that the proper functioning of temporally extended rational agency is the defining feature of some excuse defenses, including duress.

[286] *See* Alec Walen, *The Right to Cause Harm as an Alternative to Being Sacrificed for Others: An Exploration of Agent-Rights with a Special Focus on Intervening Agency*, 55 SAN DIEGO L. REV. 381, 386–87 (2018) (arguing, in relation to a trolley problem very similar to the one discussed *infra*, that the parent might have a legal duty to save their child: "[I]mposing on her son a 50% chance of being killed, so that she can save ten, seems inconsistent with the special duty she owes her son.").

[287] *See* FRANK CHAPMAN SHARP, ETHICS 44 (1928).

[288] *Id.*

[289] *Id.*

[290] *Id.*

prosecution.[291] Still, the problem poses the same sort of dilemma for the actor as does the typical duress case, namely, between exposing a family member to a threat of harm and exposing non-family members to a perhaps much greater threat. From a time-slice perspective, the rational decision in Sharp's case is to save the passengers on the train—to promote "the interests of society as a whole."[292] But Sharp makes the point, albeit only in passing, that the opposite choice also can be characterized as rational, since "[t]he welfare of society as a whole, the greater good, in short, will be most effectively furthered by everyone looking after his own family first; as it is said that the streets of Jerusalem were kept clean by everyone sweeping in front of his own door."[293] The greater good of society, in other words, will be best promoted over the long term if individuals cultivate in themselves a habit of loyalty to themselves and their family members.

In Sharp's example, of course, the actor had only seconds to decide whether to close the switch. It wouldn't be crazy to argue that reliance on habits in lieu of thoroughgoing deliberation really only is rational in cases like this, where the actor must act immediately under the press of circumstance. Habits of thought ought to give way, in other words, when the actor has hours or days or even weeks to deliberate, as Mr. Contento-Pachon did.[294] To put this another way: Maybe the efficiencies associated with habit only justify the adoption of habits broad enough in scope to control one's actions in emergencies, and no broader.[295]

This objection is not crazy. But it does appear to be wrong. James said that if we want to ensure that we do the right thing "when the hour of dire need draws nigh," we are better off cultivating broad habits rather than narrow ones.[296] In James's phraseology, we are better off "groov[ing] out" (in our brains) broad "lines of discharge" that encompass not only the critical

---

[291] *See* DRESSLER, *supra* note 62, at 100 ("Subject to a few limited exceptions, a person has no criminal law duty to act to prevent harm to another, even if she can do so at no risk to herself, and even if the person imperiled may lose her life in the absence of assistance.").

[292] SHARP, *supra* note 287, at 44–45.

[293] *Id.* at 45.

[294] United States v. Contento-Pachon, 723 F.2d 691, 693 (9th Cir. 1984); *cf.* Greenawalt, *supra* note 7, at 1912 (observing that the duress defense would apply even to someone who "remains completely cool while he steals a diamond in response to a credible threat that three strangers will be killed if he refuses").

[295] *Cf.* HARE, *supra* note 63, at 36 (observing that the utility of habits (or reaction patterns) is grounded in part on the difficulty of performing "cost-benefit analysis on the spot").

[296] JAMES, *supra* note 21, at 65.

situation but a variety of similar, less important situations.[297] James appears to be right about this. A parent's sense of duty toward her children does not disappear as soon as she has the leisure to reflect on her choices. And if it did, it is hard to imagine that her sense of duty would provide the bulwark it does against parental neglect.[298] Even broad habits, then, of the kind at work in the typical duress defense, can be rational from a temporally extended perspective.

Under this proposed account of duress, the actor's culpability would depend in part on whether his or her habit of loyalty was rational, or might have been. But the actor's culpability also would depend on a second question too. Again, habits of thought, unlike behavioral habits, do not generate behavior directly. They merely affect some of the inputs to the deliberative process. Even if the habit that generated a particular input was rational, then, we would still need figure out how the substitution of this habit-generated input affected the time-slice rationality of the actor's conduct. That is, we would still have to recalculate the time-slice rationality of the conduct after substituting the habit-generated input for the usual Hand formula risk inputs.

This sounds more complicated, and more revisionary, than it is. We began this Section by asking what the various formulations of the duress defense might mean by their references to "reasonableness." All we have really done is answer that question. It might well be that, in the usual case, we cannot do any better by way of instructing the jury on duress than by referring to the reasonable person, or to the "person of reasonable firmness," in the Model Penal Code's formulation.[299] In the usual case, moreover, since little or no evidence will be available of the actor's *actual* habits, the jury probably will wind up resolving the question of culpability on the basis of a hypothetical inquiry into the sorts of habits that might have affected the actor's conduct.

III. WHY THE PROPOSED APPROACH IS BETTER THAN THE ALTERNATIVES

Another reason for thinking that culpability depends on the desirability of the actor's habits is that no alternative account of habit's role in criminal

---

[297] *Id.* at 64; *see also* Eric A. Johnson, *Habit and Discernment in Abortion Practice: The Partial-Birth Abortion Ban Act of 2003 as Morals Legislation*, 36 RUTGERS L.J. 549, 588 (2005) ("[T]he strengthening of the habit—the widening of the path—need encompass only those situations that are *proximate* to the critical situation in the mental landscape.") (quotation and footnote omitted).

[298] *Cf.* HARE, *supra* note 63, at 137 ("The dilution of this responsibility would weaken it out of existence.").

[299] MODEL PENAL CODE § 2.09(AM. L. INST. 1985).

liability really is plausible. In this Part, I will very briefly consider three alternative approaches. The first two of these alternatives, which I will refer to as the "conscious-wrongdoing approach" and the "capacity approach," acknowledge that habits can negate culpability but deny that this culpability-negating effect depends on the habit's desirability. The third alternative, the character theory, acknowledges that culpability depends on the desirability of the actor's habits but insists that this dependence is only indirect. It insists that culpability ultimately is about moral character, rather than habit.

A. THE CONSCIOUS-WRONGDOING APPROACH

Suppose I am wrong in thinking that culpability depends on the desirability of the actor's habits. How, then, would we assess the actor's liability in cases where the actor's habits preempted some aspect of the actor's deliberations?

One alternative would be to ask simply whether the actor's habits, desirable or not, so preempted the actor's deliberations that the actor cannot be regarded as having "consciously chose[n] to do wrong."[300] Under this account, an actor is not culpable unless he or she is consciously aware, in the moment of the voluntary act, that the act is wrongful.[301] If, by virtue of a habit or anything else, the actor lacks this subjective awareness in the moment of the act, then the actor is not culpable.

To understand this approach, which I will refer to as the conscious-wrongdoing approach, we first have to distinguish two kinds of knowledge: (1) the actor's knowledge of the background facts in which the risk posed by the conduct inheres; and (2) the actor's inferential knowledge of the risk itself. The first sort of knowledge is a component of wrongdoing, as we have seen.[302] Unless the actor is aware of the background facts in which the risk inheres, the actor's creation of the risk cannot even be wrongful, much less

---

[300] Jody Armour, *Where Bias Lives in the Criminal Law and Its Processes: How Judges and Jurors Socially Construct Black Criminals*, 45 AM. J. CRIM. L. 203, 236 (2018) (describing but not endorsing this account of criminal liability).

[301] *See* Commonwealth v. Heck, 491 A.2d 212, 224 (Pa. Super. Ct. 1985), *aff'd*, 535 A.2d 575 (Pa. 1987) ("In a fundamental sense the harshness of criminal punishment is fitting only for these types of consciously inflicted wrongs, and so traditionally the criminal law has concerned itself exclusively with conscious wrongdoing.") (citation omitted); Jerome Hall, *Negligent Behavior Should Be Excluded from Penal Liability*, 63 COLUM. L. REV. 632, 634 (1963) (arguing that criminal liability is appropriate only for "voluntary harm-doing" and that voluntary harmdoing requires "at least an awareness of possible harm"); DRESSLER, *supra* note 62, at 128 (summarizing arguments for this position).

[302] *See supra* text accompanying notes 2–6.

culpable.[303] What is distinctive about the conscious-wrongdoing approach is that it would require the second kind of knowledge, too. It would impose liability only when the actor actually had put his or her background knowledge to use—actually had inferred, for example, that the conduct posed a substantial risk.

As it happens, scholars who disapprove of negligence as a basis for criminal liability sometimes make arguments very like this one.[304] They argue, roughly, that unless the actor is conscious of the risk posed by his or her conduct, the actor cannot realize that his or her conduct is wrongful.[305] And unless the actor realizes that his or her conduct is wrongful, he or she cannot be culpable.[306] On this view, then, if an actor's habits cause him or her not to be aware of the risk posed by the conduct—if, say, an actor's habit of trust causes her not to be aware of the risk posed by entrusting her car to an intoxicated friend—then the actor is not culpable.[307]

This account would not, however, merely absolve actors' whose habits caused them not to realize that their conduct posed a substantial risk. After all, being aware that one's conduct is wrongful requires a lot more than "[s]imply being aware of creating 'substantial' risks," as Jody Armour has pointed out.[308] It also requires awareness that the risk is of sufficient magnitude to outweigh the conduct's benefits.[309] Accordingly, this account also would absolve actors whose habits caused them to miscalculate the probabilities associated with negative outcomes. It even would absolve actors whose habits of valuation—or "normative contouring," in Agule's phrase—

---

[303] Johnson, *supra* note 2, at 786–88.

[304] *See* JEROME HALL, GENERAL PRINCIPLES OF CRIMINAL LAW 133–34 (2005) ("The relevant ethical principle expressed in terms of mens rea, that penal liability should be limited to the voluntary (intentional or reckless) commission of harms forbidden by penal law, represents not only the perennial view of moral culpability, but also the plain man's morality."); GLANVILLE WILLIAMS, CRIMINAL LAW: THE GENERAL PART 123 (2d ed. 1961) ("[T]he threat of punishment for negligence must pass [the negligent actor] by, because he does not realise that it is addressed to him.").

[305] DRESSLER, *supra* note 62, at 128 (explaining but not endorsing this view).

[306] *Id.*

[307] *See* HALL, *supra* note 304, at 138–-39 (acknowledging that negligent conduct may originate in the actor's "habits"—in "the 'structure of personality' which had its origin in childhood and in a vast number of subsequent experiences"—but arguing that the actor can't be held responsible for habit-generated wrongs, since his habits "are not matters of choice").

[308] Armour, *supra* note 300, at 236.

[309] *Id.*; *see also* SIMESTER, *supra* note 228, at 245 (observing that under this view (which Simester doesn't endorse) "there is no choice-based fault in choosing to take a risk that, on the facts as perceived, it is reasonable to take").

caused them to assign the wrong values to outcomes, so long as the actor, as a result of this valuation error, did not realize that the conduct was wrong.[310]

To its credit, this fully developed version of the conscious-wrongdoing account would explain why duress exculpates. On this approach, the duress defense exculpates because the actor's habit of loyalty—the actor's habit of assigning greater value to the actor's own and his or her family member's interests—causes the actor not to realize that the risks posed by his or her conduct outweigh the benefits.

To its considerable discredit, though, the conscious-wrongdoing approach would absolve just about everybody else too. Remember, the conscious-wrongdoing approach would not distinguish the effects of desirable habits—habits like trust, loyalty, and pride—from the effects of undesirable habits. Thus, for example, if a racist or homophobe were to engage in the sort of "normative contouring" that characterizes racism or homophobia and were, as a result, to fail to recognize that his or her conduct was wrongful, the racist or homophobe would be no less entitled to a defense than would someone who acts under duress. Likewise, a person who made a habit of disregarding all risks to other people and who, as a result, failed to recognize that his or her conduct was wrongful would be no less entitled to a defense than would a person whose failure to perceive a risk was rooted in a rational habit of trust.[311]

Outcomes like these would not just be troubling; they would be starkly inconsistent with existing law. Needless to say, the duress defense does not have a counterpart for racists and homophobes. Racism and homophobia aggravate wrongdoing.[312] They don't exculpate. In the majority of jurisdictions, moreover, subjective awareness by the actor that his or her

---

[310] *See* Alec Walen, *Crime, Culpability, and Moral Luck*, 29 L. & PHIL. 373, 376 (2010) (observing that "if one thinks, however mistakenly, that one has sufficient concern for others, then one will think one is doing what one ought to do"). The insanity defense does, of course, absolve some actors who are unaware that their conduct is wrong. But the insanity defense poses a question about capacity, not about actual consciousness. *See* MODEL PENAL CODE § 4.01(1) (AM. L. INST. 1985). The question posed by the defense is whether, by virtue of a mental disease or defect, the actor lacked "substantial capacity" to realize that the conduct was wrong. *Id.*

[311] At bottom, the trouble with the conscious-wrongdoing account is that even the most blameworthy wrongdoers usually don't realize that they're doing wrong. As Alec Walen has said, criminal conduct less often is the result of contemporaneous moral fault than of "a series of earlier choices . . . not to cultivate a responsible concern for others." Walen, *supra* note 310, at 383.

[312] *See, e.g.*, 720 ILL. COMP. STAT. 5/12-7.1 (2022) (imposing aggravated penalties in cases where the crime is committed "by reason of" the victim's race or sexual orientation).

conduct poses a substantial risk is not a threshold requirement of criminal liability.[313] And even those offenses that do require awareness of risk do not require proof that the actor was subjectively aware of the risk's magnitude, much less proof that the actor was aware that the risks outweighed the utility.[314]

Adherents of the conscious-wrongdoing approach do have a possible response to concerns like these. They could argue that habitual responses involve a kind of "dispositional" or "preconscious" deliberation.[315] That is, they could argue that even when habit preempts *conscious* deliberation—even when deliberation does not "enter the [the actor's] stream of consciousness"[316]—the actor *still* engages preconsciously in deliberation of the kind presupposed by the conscious-wrongdoing account.[317] This strategy finds modest support in recent claims by some psychologists that even habitual behavior is "guided by implicit goals, or needs outside of conscious awareness."[318] These "goal theorists" do not argue merely that an actor reasons about consequences when he or she exerts control over the process of habit *formation*. Instead, they argue that reasoning about consequences continues in the actor's pre-consciousness even in the moment when the actor responds to a stimulus on the basis of a settled habit.[319] For goal theorists, then, in the moment when a person fastens her seatbelt, he or she is deliberating preconsciously about how to prevent injuries.

This "dispositional awareness" strategy is a stretch. For one thing, most psychologists appear to reject goal theory in favor of the view that "habitual responses are directly cued by contexts."[320] The psychological evidence

---

[313] *See* Caleb Mason, *Framing Context, Anonymous Internet Speech, and Intent: New Uncertainty About the Constitutional Test for True Threats*, 41 SW. L. REV. 43, 104 (2011) ("Negligent homicide is found in most states' criminal codes.").

[314] *See, e.g.*, People v. Knoller, 158 P.3d 731, 742 (Cal. 2007) (holding that implied-malice murder, which requires a "high probability of death," does not require the government to prove that the actor was subjectively aware of a high probability of death).

[315] *See* Stephen P. Garvey, *What's Wrong with Involuntary Manslaughter?*, 85 TEX. L. REV. 333, 344 (2006) (distinguishing "dispositional awareness" from both unawareness and from reflective awareness).

[316] Moore & Hurd, *supra* note 21, at 192–93.

[317] *Id.* at 154. ("The acts that constitute such performances are rightly regarded as chosen despite the lack of conscious experience of the choice.").

[318] Wood, *supra* note 20, at 593 (discussing Arie W. Kruglanski & Ewa Szumowska, *Habitual Behavior is Goal-Driven*, 15 PERSP. PSYCHOL. SCI. 1256 (2020)).

[319] Kruglanski & Szumowska, *supra* note 318, at 1257 ("[A]ll human actions are driven by specific goals, although the [specific goals] do not have to be consciously represented at all times.").

[320] Wood, *supra* note 20, at 590.

suggests that "people act habitually regardless of their implicit attitudes or goals."[321] So does the neuroscientific evidence.[322] To make matters worse, psychologists have raised serious questions about whether "goal theory" even is "falsifiable"—whether it even is "open to empirical test."[323] As the Supreme Court famously said in *Daubert*, falsifiability is "what distinguishes science from other fields of human inquiry."[324] It is not even clear, then, whether the goal theorists are doing science or metaphysics.

Even if we were to suppose that habitual thoughts, feelings, and behaviors *sometimes* are informed by preconscious deliberation, it is hard to know how a factfinder would go about deciding whether, and to what extent, the actor's conduct in any particular case actually was informed by preconscious deliberation. Naturally, the goal theorists themselves have not explained how the factfinder would make this determination. Their thesis, rather, is simply that "*all* human actions" are informed by preconscious deliberation on specific goals.[325] If courts were to indulge *this* presumption, however, culpability would collapse into wrongdoing. The culpability question would just be whether the risks outweighed the benefits under the background circumstances known to the actor.

B. THE CAPACITY APPROACH

Another possible account of the relationship between habits and culpability is what I will call the capacity account. For the causal question posed by the conscious-wrongdoing account—whether the actor's habits caused him not to deliberate sufficiently on the possible consequences of the conduct—the capacity account would substitute a capacity question. It would pose the question whether, as a result of his or her habits, the actor lacked the *capacity* either to deliberate on the conduct's consequences or to bring this deliberation to bear in controlling his or her conduct.[326]

---

[321] *Id.* at 593.

[322] *See id.* at 592 ("A central finding in behavioral neuroscience is that the brain circuits activated during habit performance are separate from, but interconnected with, those associated with goal-dependent actions.").

[323] *Id.* at 590.

[324] Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 593 (1993) (quoting ERIC D. GREEN & CHARLES R. NESSON, PROBLEMS, CASES, AND MATERIALS ON EVIDENCE 645 (1983)); *see also* KARL POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY 18 (1959).

[325] Kruglanski & Szumowska, *supra* note 318, at 1257.

[326] *See* HART, *supra* note 61, at 181 (identifying as "the rationale for most of the existing excuses" "the simple idea that unless a man has the capacity and a fair opportunity or chance to adjust his behavior to the law its penalties ought not to be applied to him"); *see also id.* at 152–53 (explaining capacity theory).

It is true, of course, that our conception of wrongdoing presupposes that the actor had the capacity to do otherwise; it presupposes the "power of avoiding the evil," in Holmes's phrase.[327] The two mental states that bear on the wrongfulness of the actor's conduct—volition and knowledge of circumstances—are fundamentally about this abstract capacity. To say that the actor exercised volition is to say that she had the raw power *not* to do what she did; that the movement of her body was not, say, a muscular spasm.[328] Likewise, to say that the actor knew of the facts and circumstances in which the risk inhered is to say that she had the raw power to foresee the possible consequences, to derive the relevant probabilities.[329]

It is also true that an actor who is bereft of either capacity is not a suitable object for punishment. The traditional *M'Naghten* or cognitive variant of the insanity test absolves of liability the actor who, by virtue of a mental disease or defect, lacks the substantial capacity to appreciate the wrongfulness of his or her conduct.[330] Likewise, the "irresistible impulse" variant of the insanity defense absolves the actor who, by virtue of a mental disease or defect, lacks the substantial capacity to exercise volitional control over his or her actions.[331]

Adherents of the capacity approach to culpability would model the whole culpability inquiry on the insanity defense.[332] Under the capacity approach, the culpability question is just whether something—duress, insanity, etc.—obstructed or impaired the actor's capacity to realize that his or her conduct was wrongful or to bring this realization to bear in controlling his or her conduct.[333] In capacity theory, the focus remains on the instant of

---

[327] *See* HOLMES, *supra* note 3, at 95; *see also id.* at 54–55 ("[I]t is felt to be impolitic and unjust to make a man answerable for harm, unless he might have chosen otherwise.").

[328] *See* MODEL PENAL CODE § 2.01 (AM. L. INST. 1985) (providing that "a reflex or convulsion" does not satisfy the criminal law's "voluntary act" requirement).

[329] *See* HOLMES, *supra* note 3, at 56 ("If a consequence cannot be foreseen, it cannot be avoided."); Stephen R. Perry, *Risk, Harm, and Responsibility*, *in* PHILOSOPHICAL FOUNDATIONS OF TORT LAW 321, 344–45 (David Owen ed. 1995) (identifying foreseeability as a critical component of the capacity to avoid wrongdoing).

[330] MODEL PENAL CODE § 4.01(1); *see also* MPC COMMENTARIES, *supra* note 10, at 171–72 (characterizing pre-Code *M'Naghten* test as requiring "utter incapacity").

[331] MODEL PENAL CODE § 4.01(1) (AM. L. INST. 1985); *see also* MPC COMMENTARIES, *supra* note 10, at 171–72 (comparing Code to pre-Code version of irresistible impulse defense).

[332] *See* Brudner, *supra* note 9, at 153 ("[T]he moral involuntariness account seeks to integrate exonerating excuse with exculpation through an analogy between severely constrained choice and inoperative capacity for choice.").

[333] *See* BRINK, *supra* note 61, at 4 (arguing that "significant impairment of either [normative competence or situational control] results in an excuse").

the voluntary act.[334] But the question is not about what happened in the actor's mind in that instant. It is about what *could have* happened in her mind, given the conditions that prevailed in that instant.[335]

In its treatment of habits, the capacity approach naturally would differ both from the consciousness of wrongdoing approach and from my proposed approach. Unlike the consciousness of wrongdoing approach, the capacity approach would not ask just whether the habit preempted the actor's deliberations. It would ask instead whether the habit had deprived the actor of the *capacity* to deliberate.[336] Unlike the approach I've proposed, the capacity approach would not require the factfinder to evaluate the desirability of the habits that influenced the actor's conduct. Under the capacity approach, rather, everything that impairs the actor's exercise of his or her deliberative or volitional capacities in the moment of the conduct is presumed to be undesirable. Accordingly, under the capacity approach the exculpatory effect of a habit would depend only on the *degree* of impairment wrought by the habit—on the habit's strength or force—not on the desirability of the impairment.

The capacity approach to habit does not work. First, as applied to desirable habits, the question posed by the capacity approach is nonsensical. The question posed by the capacity approach—whether the actor *could* have overcome a particular influence—isn't just a question about what was possible or probable. It isn't just a question about what the actor *might* have done as a statistical matter.[337] Rather, in this context the question whether the actor *could* have overcome a particular influence presupposes that he or she *should* have. Recognizing the existence of the normative demand is, after all, a critical part of exercising the capacity to conform to the demand.[338] Where desirable habits are concerned, however—where the habits' efficiencies outweigh their costs in over-inclusiveness—it is just not the case that the actor *should* have overcome the habit's influence on his or her conduct.

---

[334] *Id.* at 106–09 (adopting the "ahistorical version of fair opportunity," which says that "one is excused for wrongdoing just in case one lacks sufficient normative competence or sufficient situational control at the time of action").

[335] *See* SIMESTER, *supra* note 228, at 249 (contrasting the capacity account with the "choice" account, which says that the actor's culpability inheres not in the actor's capacity to control his or her conduct but rather in the actor's "exercise of that control").

[336] *See id.*

[337] *Cf.* Moore & Hurd, *supra* note 21, at 157–60 (exploring in detail what is meant by the "capacit[y] to do other than what one did").

[338] *See* BRINK, *supra* note 61, at 54 (identifying "normative competence" as a precondition of responsibility).

Where desirable habits are concerned, acting in accordance with the habits is entirely rational, as we've seen.

The capacity approach doesn't make sense as applied to *undesirable* habits, either. Under the capacity approach, again, the exculpatory effect of habit would depend purely on the strength or force of the habit—on the *degree* to which the habit impaired the actor's capacity to deliberate or to exercise volition.[339] If the exculpatory effect of habit depends only on the habit's degree, however, then even very dangerous habits—racism, homophobia, etc.—would absolve the actor of liability so long as the habits rose to the required strength. An actor who habitually assigned no weight to the lives of Black people, for example, would be entitled to the same consideration as an actor who habitually assigned great weight to the lives of her children. Both actors would be entitled to an acquittal if, in the moment of the charged conduct, their habits affected their deliberative capacity to the required degree.

To this last point, adherents of the capacity view sometimes respond by resorting to a "tracing strategy."[340] When the actor's capacity is impaired by a dangerous habit—racism, homophobia, etc.—they would hold the actor responsible for the earlier action by which he or she *acquired* the habit.[341] Bear in mind, the tracing strategy does not expand the "time frame" to encompass both the charged conduct and the earlier moment when the person acquired the undesirable habit.[342] Rather, the tracing strategy merely repositions the usual narrow time-frame over the moment of the *earlier* choice by which the person acquired the habit. All the conditions of liability—voluntary act, knowledge of risk-engendering circumstances, causation, etc.—must be satisfied in relation to this earlier moment, then.[343]

The tracing strategy doesn't work, unfortunately.[344] It is hard to imagine a case where the government would be able to identify a particular voluntary act by which the actor had acquired the operative habit. Human beings do not

---

[339] *Id.* at 63 (arguing that "reasons-responsiveness is a matter of degree").

[340] *Id.* at 107–08, 265; *cf.* Ferzan, *supra* note 56, at 646–47 ("Alternatively, query whether the actor's acquisition of the habit suffices to serve as the voluntary act. However, for those of us that adopt a choice-based view of responsibility, whether we focus on the mens rea or the actus reus, the lack of consciousness involved in these cases is very problematic.").

[341] BRINK, *supra* note 61, at 107 ("[T]racing seems to explain how we can be responsible when we act on automatic or habitual scripts, which do not seem to be reasons-responsive.").

[342] *Cf.* Mark Kelman, *Interpretive Construction in the Substantive Criminal Law*, 33 STAN. L. REV. 591, 594 (1981) (discussing use of narrow and broad "time frames" in criminal law).

[343] *See* Moore, *supra* note 55, at 414–16 (discussing and illustrating "tracing strategy").

[344] *See id.* at 414 ("[T]he tracing strategy itself is generally a terrible argument for responsibility.").

acquire habits by just *deciding* to acquire them. They acquire them gradually and over time, usually by repeatedly "behaving in one way or the other in appropriate circumstances."[345] Even if, *per impossibile*, the government somehow were able to identify the voluntary act by which the actor had acquired the relevant habit, it is hard to imagine how the government would go about proving that the actor satisfied in that moment all the other elements of the charged offense as well. It is hard to imagine, for example, how the government would go about proving beyond a reasonable doubt that the actor, in that long-ago moment, had the requisite foresight of the consequences.[346]

Finally, and as a practical matter, it is an advantage of the habit-centered approach over the capacity approach that, except in insanity cases, it does not ask the factfinder to determine what the actor "could have done," or even what a reasonable person in the actor's position "could have done."[347] As Stephen Morse has argued, questions about whether the actor "couldn't help it"—about whether the actor could or couldn't have overcome "internal duress" of the sort generated by habits—are conceptually unclear and perhaps unanswerable.[348] The very different questions posed by the habit-centered test of culpability—the question what habits played a role in the actor's conduct, or might have, and the question whether these habits are justified—are not easy. But unlike questions about what the actor could have done, they are meaningful.

C. THE CHARACTER APPROACH

A third alternative to the habit-centered theory is character theory, which says that culpability is about the actor's moral character. In character

---

[345] ARISTOTLE, NICOMACHEAN ETHICS Bk. II, ch. 1, 1103b (W.D. Ross, transl.) *in* THE BASIC WORKS OF ARISTOTLE 953 (Richard McKeon, ed. 1941); *see also* JAMES, *supra* note 22, at 12 ("*Continuity* of training is the great means of making the nervous system infallibly right."); Moore & Hurd, *supra* note 21, at 186 ("From an early age, each of us has drilled into us a set of rules which our parents hope will become 'automatic' or routine . . . . As we become older, additional rules are added through social osmosis.").

[346] *Cf.* MPC COMMENTARIES, *supra* note 10, at 359 (acknowledging, in relation to voluntary intoxication, "the impressive difficulties posed in litigating the foresight of any particular actor at the time when he imbibes").

[347] This represents a departure from Model Penal Code § 2.09, for example, which frames the duress question in terms of capacity. Under § 2.09, the question is not whether a reasonable person would have engaged in the proscribed conduct but, rather, whether a person of reasonable firmness would have been "unable to resist."

[348] *See* Stephen J. Morse, *Uncontrollable Urges and Irrational People*, 88 VA. L. REV. 1025, 1056, 1059 (2002).

theory, the question posed by the culpability component is whether the actor's wrongdoing provides a basis for inferring that he or she has a particular sort of character defect—"insensitivity to the interests of others," for example.[349] In character theory, as A.P. Simester has said, judgments of blame in criminal law "start with wrongdoing, and end by tracing that action to a moral vice in the actor."[350]

To its credit, character theory can explain what capacity theory cannot, namely, why the law distinguishes actions generated by desirable habits—actions performed under duress, for example—from actions generated by undesirable habits, like racism and homophobia. The difference between these two cases, according to character theory, is that in one case but not the other, it is possible to trace the actor's wrongdoing to moral vice.[351] When the actor engages in wrongdoing under duress, the existence of duress blocks the usual inference from wrongdoing to vice.[352] In effect, the duress provides an alternative, non-vicious explanation for the wrongdoing, specifically, the desirable habit of loyalty to oneself and one's family members. In contrast, the existence of racial animus, say, obviously doesn't block the usual inference from wrongdoing to vice, regardless of how it affects the actor's self-control. On the contrary, the undesirable habit itself, the actor's racism, provides a basis for inferring that the actor is culpably insensitive to the interests of others.

It would be easy to get the impression that habit theory is really just a slight variant of character theory. After all, "[t]he distinction between habit and character is a difficult one to make," at least in the law of evidence.[353] And indeed, some of the propensities that I have identified as habits—loyalty, trust, pride, etc.—could easily be described as character traits. So the

---

[349] Johnson, *supra* note 238, at 968–72 (arguing that the phrase "insensitivity to the interests of others," as used in the MPC COMMENTARIES, *supra* note 10, at 243, reflects the drafters' recognition that negligence is culpable only when it betrays a defect in the actor's character or disposition); *see also* DRESSLER, *supra* note 62, at 203–04 (discussing character theories); SIMESTER, *supra* note 228, at 250 (identifying lack of "sufficient concern" for others as the sort of "moral vice" that is the concern of criminal law).

[350] SIMESTER, *supra* note 228, at 253; *see also* Herbert Wechsler & Jerome Michael, *A Rationale for the Law of Homicide: Part II*, 37 COLUM. L. REV. 1261, 1275 (1937) (explaining that excusing conditions in effect prevent the factfinder from "relat[ing] the evil of the act to the actor's desires in any way that bears significantly on his character").

[351] *See* DRESSLER, *supra* note 62, at 203–04 ("[C]haracter theorists argue that excuses should be recognized in the law in those circumstances in which bad character cannot be inferred from the offender's wrongful conduct.").

[352] *See* FLETCHER, *supra* note 228, at 799 ("The distinguishing feature of excusing conditions is that they preclude an inference from the act to the actor's character.").

[353] Figueiredo v. Hamill, 431 N.E.2d 231, 232 (Mass. 1982) (citations omitted).

question arises: In the end, is there really any difference between identifying culpability with undesirable habits, as I have done, and identifying culpability with moral vice, as the character theorists would do?

There is a big difference, I think. The question posed by character theory is not whether the actor's wrongdoing can be traced to just any vice—to just any dangerous or undesirable habit. Rather, the question posed by character theory is whether the actor's wrongdoing can be traced to a kind of foundational vice—"insensitivity to the interests of others," for example.[354] This foundational vice can make itself felt in undesirable habits, just as it can make itself felt in wrongful conduct. But the foundational vice is not the same as bad habits. Nor can we infer merely from the fact that an actor has bad habits that the person is afflicted by the foundational vice.

The reason we cannot draw this inference is that variations in habit among individuals sometimes are attributable to "differences in circumstances" or even to differences in genetic makeup,[355] rather than to differences in sensitivity to others' interests. Suppose, for example, that a particular individual suffers from a genetic predisposition to violence, or a genetic predisposition to exceptional timidity.[356] Such an individual might try to compensate for his or her predisposition by cultivating habits that would be unreasonable in the average member of the community, and so might wind up adopting habits that in the average member of the community would reflect insufficient concern for others.[357] To infer from an actor's habits that he or she is afflicted by the foundational vice, then, the factfinder would need to take into account the actor's life circumstances, the actor's upbringing, the actor's heredity, etc.

So habit theory differs from character theory. Under habit theory, the culpability component is satisfied when the actor's wrongdoing is attributable to habits of thought, feeling, or behavior that would be unreasonable in an ordinary member of the community. Under character theory, by contrast, the culpability requirement is satisfied only when it is

---

354 Johnson, *supra* note 238 (quoting MPC COMMENTARIES *supra* note 10, at 243).

355 BRATMAN, *supra* note 13, at 70 ("[H]abits that it is reasonable for me to have may not be reasonable for you to have.").

356 *See* Hannah L. Bedard, *The Potential for Bioprediction in Criminal Law*, 18 COLUM. SCI. & TECH. L. REV. 268, 288–89 (2017) (discussing State v. Waldroup, No. E2012-00758-CCA-RMCD, 2013 WL 4279213 (Tenn. Crim. App. Aug. 15, 2013), where a defendant charged with murder presented expert testimony that his MAOA genotype, together with a childhood history of abuse, predisposed him to violent actions).

357 *See* BRATMAN, *supra* note 13, at 70 (explaining that in circumstances like these, "we could say that the Devil has made it rational for you to try to develop unreasonable habits of reconsideration").

possible to infer from the all the evidence—the wrongful act, the habits that generated it, and the actor's circumstances, heredity, etc.—that the actor lacks sufficient concern for the interests of others.

Of the two theories, only habit theory can seriously claim to reflect the existing law. Again, we cannot know whether the actor is afflicted by the foundational vice—lack of concern for the interests of others—without knowing about the actor's upbringing, heredity, etc. But the existing law of culpability forecloses inquiry into matters like upbringing and heredity.[358] The Model Penal Code commentary, for example, says that "the heredity, intelligence or temperament of the actor would not be held material in judging negligence, and could not be without depriving the criterion of all its objectivity."[359] Likewise, in explaining the duress defense, the Code commentary says that an actor's "peculiar disabilities"—exceptional timidity, for example—have no bearing on the defense: "To make liability depend upon the fortitude of any given actor would be no less impractical or otherwise impolitic than to permit it to depend upon such other variables as intelligence or clarity of judgment, suggestibility or moral insight."[360]

The existing law surely gets this right. As Holmes said, the criminal law's "immediate object and task [is to] establish a general standard, or at least general negative limits, of conduct for the community . . . ."[361] This is no less true, ultimately, with respect to culpability than with respect to wrongdoing. Just as the requirement of wrongdoing aims to establish a general, external standard of *conduct*—of time-slice rationality—the requirement of culpability ought to aim to establish a general, external standard of temporally extended rationality. It ought to educate us about "what habits (and so on) to encourage and develop in ourselves and others over the long run."[362] If this is our objective in developing standards for temporally extended rational agency, then we "want standards that are shaped by a concern for normal cases, given general cultural and

---

[358] *See* DRESSLER, *supra* note 62, at 129; *id.* at 300.

[359] MPC COMMENTARIES, *supra* note 10, at 242.

[360] *Id.* at 374.

[361] Commonwealth v. Pierce, 138 Mass. 165, 176 (1884); *see also* HOLMES, *supra* note 3, at 50 ("[W]hen we are dealing with that part of the law which aims more directly than any other at establishing standards of conduct, we should expect there more than elsewhere to find that the tests of liability are external . . . .").

[362] BRATMAN, *supra* note 13, at 70.

psychological constraints," as Bratman has said.[363] We do not want standards that hinge on the actor's heredity, intelligence, etc.

CONCLUSION

Criminal law's wrongdoing component is anything but mysterious, at least for adherents of the tradition that extends from Holmes through Wechsler and the Model Penal Code.[364] Conduct is wrongful when, "under the circumstances known to [the actor]," the risks posed by the conduct are greater than the conduct's benefits.[365] This risk-based conception of wrongdoing is more than just elegant. It has important practical advantages too. For one thing, it provides a clear alternative to legal moralism—to the view that wrongdoing is just anything the majority disapproves of.[366] For another, it explains why knowledge of the facts in which the risks inhere is an absolutely fundamental element of criminal liability.[367]

For adherents of this same tradition, though, culpability represents something of an embarrassment. Holmes was skeptical about whether the criminal law required anything more than wrongdoing—about whether "the actual degree of personal guilt involved in any particular transgression . . . is an element at all."[368] By contrast, Wechsler realized that criminal liability

---

[363] *Id.*; *see also* MODEL PENAL CODE § 210.3 cmt. 5(b) at 71 (explaining that when the law "evaluat[es] the abnormal individual on his own terms . . . . [i]t blurs the law's message that there are certain minimal standards of conduct to which every member of society must conform").

[364] *See* Johnson, *supra* note 4, at 515–27 (tracing the Model Penal Code's risk-centered conception of wrongdoing back through Herbert Wechsler's scholarship and ultimately to Holmes's *The Common Law*).

[365] MODEL PENAL CODE § 2.02(2)(c), (d) (AM. L. INST. 1985); *see also* Herbert Wechsler & Jerome Michael, *A Rationale for the Law of Homicide: Part I*, 37 COLUM. L. REV. 701, 710 n.31 (1937) (citing Holmes for the proposition that the actor's unawareness of the consequences of his actions is immaterial "if under the circumstances known to him" the consequences are obvious).

[366] *See* Lawrence v. Texas, 539 U.S. 558, 577 (2003) (embracing the view that "the fact that the governing majority in a State has traditionally viewed a particular practice as immoral is not a sufficient reason for upholding a law prohibiting the practice").

[367] *See* Rehaif v. United States, 139 S. Ct. 2191, 2196 (2019) ("The cases in which we have emphasized scienter's importance in separating wrongful from innocent acts are legion."); Johnson, *supra* note 2, at 786–88 (explaining why the mens rea presumption is rooted in the wrongdoing component of criminal liability, rather than the culpability component).

[368] HOLMES, *supra* note 3, at 49; *see also id.* at 75 ("[T]he *mens rea*, or actual wickedness of the party, is wholly unnecessary, and all reference to the state of his consciousness is misleading if it means anything more than that the circumstances in connection with which the tendency of his act is judged are the circumstances known to him.").

required more than wrongdoing but struggled to say exactly what it required.[369] The Model Penal Code reflects this uncertainty. Of culpability, the Code's comments say simply, and inadequately, that "the jury must evaluate the [actor's] conduct and determine whether it should be condemned."[370] It is hard to see how a "test" like this can be squared with the beyond a reasonable doubt standard, or with the requirement that criminal law be knowable.[371]

The answer to the riddle of culpability begins, as we have seen, with the recognition that rationality is irreducibly twofold. The kind of rationality that we associate with desirable habits—with habits whose risks are outweighed by their benefits—is not captured when we analyze the time-slice rationality of specific acts. As a consequence, even wholly rational habits can generate acts that are, from a time-slice perspective, irrational, or wrongful. From this insight, it is possible to construct a conception of culpability that not only is largely consistent with existing law—with the law of duress, for example, and of proximate cause—but also is no less elegant and determinate than our fundamental conception of wrongdoing.

---

[369] *See* Wechsler & Michael, *supra* note 350, at 1275 ("We cannot avoid asking why the actor did not pay attention when so much turned out to be at stake.").

[370] MODEL PENAL CODE, § 2.02 cmt. 3 at 125–126 (AM. L. INST., Tentative Draft No. 4, 1955).

[371] *See* Burrage v. United States, 571 U.S. 204, 217–18 (2014) (suggesting that "uncertainty" of the kind associated with the "contribution" test of cause-in-fact "cannot be squared with the beyond-a-reasonable-doubt standard applicable in criminal trials or with the need to express criminal laws in terms ordinary persons can comprehend").