

FREE AGENTS VS. DETERMINED: NEUROSCIENCE AND THE LAW OF PERSONAL RESPONSIBILITY

Susan D. Carle & Tara L. White***

TABLE OF CONTENTS

INTRODUCTION	358
I. METAPHYSICS	361
II. NEUROSCIENCE.....	363
A. <i>Sapolsky's Argument in Determined</i>	363
B. <i>Mitchell's Free Agents</i>	371
C. <i>Spatial and Temporal Resolution of Human Agency: The Agentic Envelope</i>	373
III. LAW.....	383
A. <i>The Basics of Corrective Justice</i>	383
B. <i>Rights and Responsibilities as a Package Deal</i>	385
C. <i>Reading Neuroscience and Law Together</i>	387
D. <i>Utility, Aperture, and Applications of the Agentic Envelope in Law</i>	388
CONCLUSION.....	390

* Professor of Law, American University Washington College of Law (WCL); co-host of the Lobes & Robes Podcast sponsored by the American University Center for Neuroscience and Behavior. The author wishes to thank Dr. Terry Davidson, Professors Andrew Ferguson and Ira Robins, Dr. Diane Carle, and Dr. Sacha Greer for sharing their expertise and feedback on earlier drafts. Sydney Robson and Olivia Lubarsky (both WCL 2025) provided excellent research assistance. Both authors thank Dr. Francis Shen and the participants in the Harvard Law & Neuroscience Workshop for the opportunity to present this Article as a work in progress and for excellent feedback during our presentation, and Drs. Kevin Mitchell and Robert Sapolsky for feedback on an earlier draft.

** Assistant Professor of Behavioral and Social Sciences, Brown University. Dr. White's lab investigates the neuroscience of human agency as well as drug effects on the brain. She has also pioneered the emerging field of Dignity Neuroscience, which explores the intersections between neuroscience and international human rights law. *See* note 40 *infra*.

INTRODUCTION

The new millennium has seen the rise of an exciting new interdisciplinary field, aptly coined “neurolaw,” which explores potential cross-fertilization across neuroscience and law.¹ Participants have been especially interested in exploring how advancements in neuroscience can contribute to law’s understanding of criminal behavior, mental states, human agency, and personal responsibility.² Offering important contributions to this ongoing inquiry, two prominent neuroscientists, Robert M. Sapolsky and Kevin J. Mitchell, have recently published engaging books, accessible to non-expert audiences, that address questions of human agency from contrasting perspectives. Sapolsky is a prominent voice insisting that neuroscience proves determinism and must force a revolution in criminal law because no one can rightly be considered to be “responsible” for their acts, whether good or bad.³ Mitchell espouses a very different view, expressing discomfort with determinism and describing living organisms as profoundly agentic.⁴ The differences in their views probably arise, in part, from differences in their interests and training,⁵ but we will not venture into those questions here, except to say that the neuroscientist half of our writing duo hails from yet another subfield studying the neurobiology of personality and emotions and focuses her research on how concepts of agency contribute to human well-being.⁶ We present that alternative viewpoint in Section II.C below.

Throughout, we present our dialog across the fields of neuroscience and law.⁷ A neurolaw interdisciplinary collaboration led us to each other

1. See Francis X. Shen, *Toward a Definition of “Neurolaw”*, 15 U. ST. THOMAS J.L. & PUB. POL’Y 174, 174 (2021).

2. See, e.g., Peter A. Alces & Robert M. Sapolsky, *Nohwere*, 63 WM. & MARY L. REV. 1079, 1086–97 (2022) (discussing neuroscience in relation to concepts of “human agency” in contracts, torts, and criminal law).

3. ROBERT M. SAPOLSKY, DETERMINED: A SCIENCE OF LIFE WITHOUT FREE WILL 43–45 (2023).

4. KEVIN J. MITCHELL, FREE AGENTS: HOW EVOLUTION GAVE US FREE WILL ix–xii (2023) [hereinafter MITCHELL, *Free Agents*].

5. To trace neuroscientists’ intellectual “family trees,” see NEUROTREE, <https://neurotree.org/neurotree/>.

6. See, e.g., Tara L. White & Meghan A. Gonsalves, *Dignity Neuroscience: Universal Rights Are Rooted in Human Brain Science*, 1505 ANNALS N.Y. ACAD. SCI. (SPECIAL ISSUE) 40 (2021); Tara L. White et al., *Anger, Agency, Risk and Action: A Neurobehavioral Model with Proof-of-Concept in Healthy Young Adults*, 14 FRONTIERS IN PSYCH. 1, 1 (2023).

7. The law professor half of our writing duo has written for many years about conceptions of human agency in law. See, e.g., Susan D. Carle, *Why the U.S. Founders’ Conceptions of Human Agency Matter Today: The Example of Senate Malapportionment*, 9 TEX. A&M L. REV. 533, 533–97 (2022) [hereinafter Carle, *Founders’ Conceptions*]; Susan D.

and to this project,⁸ through which we not only hope to explore how the neuroscience discoveries Sapolsky and Mitchell discuss affect traditional legal assumptions about personal responsibility, but also to model the generative potential of interdisciplinary dialogue.

On the subject of personal responsibility and human agency, the relevant literatures—far too vast to summarize—span the fields of metaphysics (defined as the non-empirical study of the nature of being),⁹ neuroscience, and law. This very vastness can obscure an important point we want to emphasize: the debate often referred to as the free will/determinism debate takes place around *different questions* in the three fields of (1) metaphysics, (2) neuroscience, and (3) law. In each of these fields, even the basic terms of the debate differ, which can lead to confusion. Although each field has much to offer the others, to assume that one field's answers dispose of the core questions in any of the other fields is often to commit a translation error. To be sure, debates in metaphysics, neuroscience, and law often spill into each other, sometimes very much to the good, as we will show below. But they can also lead to dramatic misunderstandings, as we also discuss. Indeed, even within neuroscience, disagreements about fundamental models, such as the use of mechanistic metaphors of the brain as a machine, promise to continue far into the future, as attested by the very disagreements between Sapolsky and Mitchell that we discuss below.

As we will argue, Sapolsky's very points about the unpredictability of complex systems refute claims that the free will/determinism debate will soon (if ever) be solved by science. Science depends on prediction, measurement, and disproving null hypotheses to achieve new knowledge; it has no way to measure or prove whether determinism is true. In turn, Mitchell's discussion of how living organisms act for "reasons" enriches interdisciplinary discourse by reinforcing what neurolaw scholars, such as Professor Stephen Morse, have been arguing for some time, which is that what all law means by "free will" is that human beings act for

Carle, *Conceptions of Agency in Social Movement Scholarship: Mack on African American Civil Rights Lawyers*, 39 LAW & SOC. INQUIRY 522, 522 (2014); Susan D. Carle, *Structure and Integrity*, 93 CORNELL L. REV. 1311, 1330 (2008); Susan D. Carle, *Theorizing Agency*, 52 AM. U. L. REV. 307, 309 (2005).

8. Dr. Tara White appeared on the Lobes and Robes podcast, on which Susan Carle is a co-host. See Lobes and Robes, *What Is Dignity Neuroscience?*, AM. U. (May 11, 2023), <https://www.american.edu/research/what-is-dignity-neuroscience.cfm>.

9. See *Metaphysics*, OXFORD ENG. DICTIONARY, https://www.oed.com/dictionary/metaphysics_n?tab=meaning_and_use (last visited Feb. 16, 2024) (defining metaphysics as "[t]he branch of philosophy that deals with the first principles of things or reality, including questions about [abstract concepts such as] being, substance, time and space, causation, change, and identity" or "abstract [theory] talk with no basis in or relevance to reality"). In other words, metaphysics deals with questions apart from empirical inquiry though science.

reasons.¹⁰ This is what the law of personal responsibility requires for legitimacy. We also offer another take on the neuroscience of human agency that further enriches the discussion, introducing the concept of the “agentic envelope,” and highlighting the empirical research on the importance of this concept to human flourishing and well-being.

Our Article unfolds as follows. In Part I, “Metaphysics,” we sketch a few points about the massive free will/determinism debate in philosophy that are relevant to our discussion. In Part II, “Neuroscience,” we briefly critique the two books under review. Section II.A commends the strengths of Sapolsky’s work in bringing complex bio-scientific theory to non-experts but questions what we view as too quick a move from his expertise in science to the very different knowledge domain of law. Section II.B shows the consilience¹¹ between Mitchell’s thesis in *Free Agents* and the point neurolaw scholars such as Morse have been making—namely, that as long as individuals are capable of acting for “reasons,” the law of personal responsibility stands regardless of whether all actions have prior causes. We also offer a third perspective from neuroscience different from both Sapolsky’s and Mitchell’s, based in research specifically aimed at understanding human agency from a neural perspective. We propose that different theories work better or worse depending on what issues they are centrally concerned with investigating, and that importing theoretical assumptions intended for one set of questions to address very different questions can lead to “translation errors” and a real risk of making problems worse rather than better.

Finally, Part III, “Law,” sketches the fundamentals of corrective justice and law’s double-sided concepts of individual rights and responsibility. This explanation helps shed additional light on our concerns about Sapolsky’s proposal to discard wholesale all ideas of personal responsibility in law. Personal responsibility is the flip side of the notion that individuals have rights to self-determination, and those rights are key to important civil liberties protections against undue government control. We end with short concluding observations.

10. See Stephen J. Morse, *Internal and External Challenges to Culpability*, 53 ARIZ. ST. L.J. 617, 618 (2021) [hereinafter Morse, *Challenges to Culpability*]; see also Stephen J. Morse, *Neuroscience, Free Will, and Criminal Responsibility*, in FREE WILL AND THE BRAIN: NEUROSCIENTIFIC, PHILOSOPHICAL, AND LEGAL PERSPECTIVES 251 *passim* (Walter Glannon ed., 2015) [hereinafter Morse, *Free Will*].

11. See EDWARD O. WILSON, CONSILIENCE: THE UNITY OF KNOWLEDGE 8 (1998) (defining consilience as “a ‘jumping together’ of knowledge by the linking of facts and fact-based theory across disciplines to create a common groundwork of explanation”).

I. METAPHYSICS

The free will/determinism debate is rife with terminology that can lead to confusion. In metaphysics, for example, the term “determinism” can refer to various propositions, from the idea that nothing other than molecules and neurons have causal force in explaining human behavior¹²—a proposition with which neither Sapolsky or Mitchell agrees—to the idea that a slew of influences, including not only neuronal workings but also environment, heredity, personality variation, and much more, cause observed human conduct. The two authors locate themselves differently in these metaphysical debates. Sapolsky refers to himself as believing in “hard incompatibilism,”¹³ by which he means both that he believes in determinism and that he does not view determinism as compatible with any theory of “free will.” This perspective can be contrasted to “compatibilism,” which argues that there is room for at least some free will even within a determinist perspective.¹⁴ Mitchell, on the other hand, states that he is dissatisfied with both “strictly determinist” and “compatibilist” positions;¹⁵ as we understand his perspective, he is closer to incompatibilist nondeterminism.¹⁶ We will not venture further into these very large debates about various theories of determinism and indeterminism, other than to point out that they very much remain at the level of metaphysics and are not subject to being answered empirically at present.¹⁷ Science can neither prove nor disprove

12. For a discussion and critique of this position, see generally MICHAEL S. MOORE, *MECHANICAL CHOICES: THE RESPONSIBILITY OF THE HUMAN MACHINE* 207–45 (2020).

13. SAPOLSKY, *supra* note 3, at 10 n.† (stating that he views this classification as essentially synonymous with “hard determinism”).

14. On compatibilism, see generally Michael McKenna & D. Justin Coates, *Compatibilism*, STAN. ENCYC. OF PHIL., <https://plato.stanford.edu/entries/compatibilism/> (Feb. 1, 2024).

15. MITCHELL, *Free Agents*, *supra* note 4, at x.

16. On incompatibilist, nondeterministic theories of free will, see Randolph Clarke et al., *Incompatibilist (Nondeterministic) Theories of Free Will*, STAN. ENCYC. OF PHIL., <https://plato.stanford.edu/entries/incompatibilism-theories/> (Aug. 18, 2021).

17. Similarly inaccessible to empirical inquiry at this juncture are questions as to whether artificial intelligence (“AI”) exhibits agency. We doubt those questions will be subject to resolution soon, for reasons Mitchell and Sapolsky both discuss. See MITCHELL, *Free Agents*, *supra* note 4, at 254 (“[W]e turn isolated elements of knowledge into a more general *understanding* of how the world works, something that artificial intelligence still struggles to do.”); *id.* at 297 (“Understanding causality can’t come from passive observation [c]ausal knowledge thus comes from causal intervention in the world. . . . The implication is that artificial general intelligence will not arise in systems that only passively receive data.”). Sapolsky discusses AI only in one footnote, noting the tendencies of machine learning algorithms to come up with solutions that are “bizarre” or focus on the wrong data, such as an AI that seemed to learn to diagnose melanomas but, it turned out, had learned to recognize that lesions photographed with rulers beside them were likely to be malignant.

determinism, for reasons Sapolsky in fact highlights, as we discuss below. Science's *operating assumption* of a mechanistically caused universe is part of its method of inquiry based on the goal of proving or disproving causal relations. But when science seeks to impose those assumptions on different knowledge domains, it moves outside its realm of expertise and has no special authority to pronounce how human affairs should be organized. From a consequentialist perspective, the operating—and unproved—assumptions of science might lead to results in the realm of political and legal ideas that would harm rather than enhance human well-being. Evidence matters on those questions, and it is not compatible with science's own orientation to assert what steps would or would not improve human well-being without reference to what the data show. We therefore briefly survey some relevant data; doing so does nothing to reduce our skepticism about the usefulness of blanket hard determinism to cancel law's use of ideas of personal responsibility in imposing liability for harms caused to others.

Likewise, even in metaphysics, what “free will” means is a highly contested question. “Libertarian” versions of the theory (which have no relationship to political libertarianism) propose that humans decide on their actions with *no* constraints; this seems implausible, as most philosophers today agree.¹⁸ More moderate theories posit that, for sure, humans act within many constraints, but still have some “free will” in choosing actions.¹⁹ Those who are in this camp advance many complex variations on such arguments, but we will not worry about all that here since our very point is that law does not require “free will” in any of those strong senses. Still another group of metaphysicians call themselves “free will skeptics”; they correctly note that there are positive social and legal consequences to being skeptical about whether human beings act with “free will.”²⁰ Even the two authors of this Article discovered through discussion that we may hold somewhat different views on some of these metaphysical matters. But that fact has not gotten in the way of our analysis here because we both agree with the proposition we will use as our operating assumption. This is the proposition that law uses, which makes the least bold assertion of all.

SAPOLSKY, *supra* note 3, at 162 n†. Thus, it appears, neither author sees AI's capacities, independent of human supervision and judgment, as having yet advanced to anywhere near the competence levels at which serious questions about AI agency would begin to emerge.

18. MOORE, *supra* note 12, at 268–69.

19. See GREGG CARUSO, REJECTING RETRIBUTIVISM: FREE WILL, PUNISHMENT, AND CRIMINAL JUSTICE 37 (2021).

20. See *generally id.*; DERK PEREBOOM, LIVING WITHOUT FREE WILL xx (2001). Our thanks to Sacha Greer for this point and these references.

The assumption law needs to function is that human beings act with “agency”; in other words, law has an agentic orientation. Agency is typically defined along the lines that most humans generally can act with minimal rationality to evaluate options, choose among them based on reasons, and then have capacity, more or less, to execute these decisions through action, as we discuss below. As others have pointed out, this definition is tautological in that it posits that law needs most individuals to be able to act for minimally rational reasons and that therefore a theory that posits that most individuals are able to act for minimally rational reasons is correct.²¹ We see this point, and we are not arguing that this definition is the “right” one since that proposition cannot be proved. What we are arguing is that, for pragmatic purposes, it is the one that fits best with the organization of human affairs; in other words, it is the best definition as a matter of political economy. It is a version of free will that law needs and uses, and it is a definition that allows law to be more humane toward individuals who cannot act based on reasons and with minimal rationality. In that sense, it creates avenues for law reform that could protect the persons Sapolsky is the most adamant about protecting, especially those whose life circumstances or unusual brain conditions most clearly belie law’s assumptions about agentic action. We discuss this further below.

II. NEUROSCIENCE

A. *Sapolsky’s Argument in Determined*

Sapolsky’s central thesis in *Determined* is that it is time to throw out the more timid ideas that *some* people *sometimes* have less self-control and cannot choose their actions (i.e., the human agency assumption we just proposed above) and replace those ideas with the blanket proposition that no one “chooses” their actions in any sense that should be relevant to law.²² As he argues, “there can be no such thing as blame, and . . . punishment as retribution is indefensible.”²³ By the same token, Sapolsky argues, no one has earned the right to be admired or treated better than others either.²⁴ Sapolsky asserts that human behavior is

21. Memorandum from Sacha Greer to authors (Sept. 14, 2023) (on file with authors).

22. SAPOLSKY, *supra* note 3, at 4–5.

23. *Id.* at 5.

24. *Id.*

caused by prior causes “all the way down,” as he puts it,²⁵ and proposes that neuroscience therefore requires a radical alteration of the basic rules of criminal responsibility.²⁶

Sapolsky provides a very basic proposal as to how law should address behavior that harms others, based on his fervent view that all traditional rules based on personal responsibility should be discarded. What he calls for is instituting a “quarantine” system—a proposal we find ill-considered, as we will discuss shortly. Here, he draws on the work of free will skeptic Derk Pereboom,²⁷ who argues for using quarantining systems *after* persons are convicted of crimes. It appears that Sapolsky would dispense with this stage of criminal proceedings and go straight to dangerousness assessments and quarantine. This, we argue below, seems a bad idea, especially in light of the empirical evidence on the use of predictive systems in criminal law, which indicates that the results are no more just on Sapolsky’s very terms, which focus on the morally outrageous effects of social class and race in the criminal justice system (about which we agree). But first we briefly highlight the outstanding strengths of his book, which lie in its accessible explanations of aspects of contemporary scientific knowledge that can be of significant use in legal theory.

Sapolsky starts by debunking the notion that, if one starts with something simple in biology and then ends up with something complex

25. In a communication with the authors commenting on an earlier draft of this Article, Dr. Sapolsky wrote that:

[I]t is easy to perceive the massive cable of causality linking, say, a fetal brain that was [exposed to] alcohol to anti-social behavior in the here and now. What is . . . harder is what contemporary behavioral biology teaches us, which is that causal links from back when are usually no thicker than a silk thread, miniscule. . . But that when you put them all together, they are as thick and causal as the massive cable of [genetic] or [acquired brain disorders].

Email from Robert M. Sapolsky, Professor, Stan. Univ., to Susan D. Carle, Professor of L., American Univ. Wash. College of L., and Tara L. White, Assistant Professor, Brown Univ. (Jan. 24, 2024, 3:05 PM EST) (on file with authors). This fascinating observation regarding cables and threads of causality suggests that concepts of causal pathways, both within and across individual lives, and the agentic envelope (i.e., what we have control over in the here and now, to which both rights and responsibilities refer) are not identical and can be disentangled. His insight suggests that a “wide causal pathway” is not intrinsically at odds with the concept of an agentic envelope; instead, one’s foundational assumptions or world view as to whether one has free will, choice, or agency rest on whether one starts from an initial premise that the universe and human life is entirely determined, as Sapolsky claims, or is instead intrinsically indeterminant, as Mitchell argues. The concept of the agentic envelope can span both views, but its implications for free will, choice, agency and law rest on one’s underlying view, especially as it to the temporal scale of human life, as Mitchell and we discuss.

26. SAPOLSKY, *supra* note 3, at 2–3.

27. *Id.* at 349–51.

and unpredictable, that suggests that free will must have “just happened.”²⁸ Sapolsky explains why this proposition is incorrect. Chaos theory and current thinking about complex emergent systems establish that applying even a simple, consistent rule for change over multiple generations of events produces wildly interesting, completely unpredictable mature states, such as complex patterns on seashells or, conversely, extinction. There is no way to predict which outcome will occur except by working through the exercise of applying the rule to one generation of change and then the next.²⁹ But even though results are unpredictable, they are determined: one rule causes only one possible change in each generation, even though the end result cannot be predicted and observing the end state gives one no predictive power over what the starting state was. Put otherwise, what Sapolsky calls “linear, additive predictability” cannot explain complex things like living organisms; “[a] cell, a brain, a person, a society, was more like the chaoticism of a cloud than the reductionism of a watch.”³⁰ Determinism and predictability are very different just as ontology and epistemology are different: one is about what is actually happening while the other is about what we can know.³¹

This discussion can be important to legal theorists, some of whom are already writing about chaos theory and complex emergent systems.³² But the same point also applies to the mechanistic, deterministic assumptions under which many neuroscientists operate. The idea that all observed phenomena have prior causes is a foundational assumption in science, but it cannot be proven. There is no measurement one can take—no “reductive linear predictability”—that explains observed end states. Thus, determinism could be true, but it also could be false; the techniques of science, which depend on the ability to measure cause and effect, and provide “proof by disproof,”³³ cannot at present tell us that. We are back to metaphysical debate. There is nothing wrong with that, but it does not seem to us to provide a very solid basis for overthrowing the centuries of development of civil liberties protections that are built into criminal law.

Sapolsky notes that, while neuroscientists can predict with a reasonable degree of certainty that a person with extensive frontal

28. *Id.* at 127.

29. *Id.* at 138–40.

30. *Id.* at 145.

31. *Id.* at 148.

32. See, e.g., Daniel A. Farber, *Probabilities Behaving Badly: Complexity Theory and Environmental Uncertainty*, 37 U.C. DAVIS L. REV. 145, 146 (2003).

33. See Stephen J. Ceci & Urie Bronfenbrenner, *On the Demise of Everyday Memory*, 46 AM. PSYCH. 27, 28 (1991) (discussing the “proof by disproof” concept).

cortical damage will exhibit socially inappropriate behavior, they cannot predict with any reliability whether that person will behave as an “impulsive murderer” or be “rude to a dinner host.”³⁴ But neuroscience’s lack of predictive power, Sapolsky points out, does not mean that brain damage did not “cause” the resulting behavior.³⁵ This is a lesson from chaoticism and complex emergent systems theory: it may never be the case that behavior can be entirely predicted, but that does not mean that antisocial behavior is not caused by anomalies of the brain.³⁶

So far so good, but Sapolsky then asserts his central thesis, which is that “subtract[ing] responsibility” from our view of behavior will make “the world a better place.”³⁷ In light of the very lessons he so convincingly teaches in his book, we are surprised at his confidence in this prediction. From a civil liberties perspective, alarm bells go off about the apparent downsides of Sapolsky’s proposal. He asks the reader to imagine what the world would look like if “trials were abolished, replaced by mere investigation[s] to figure out who actually carried out some act, and with what state of mind.”³⁸ (Is that not that already what trials are for?) His picture: “[n]o prisons, no prisoners.”³⁹ How, then, will the public be protected from persons whose behavior indicates that they are a risk to public safety? Through “quarantine”⁴⁰—what society should do is constrain a person who is dangerous to those around them “the absolute minimal amount needed to protect everyone, and not an inch more.”⁴¹ The “more danger is posed in the future, the more constraint[]” will be needed.⁴²

What standard would Sapolsky use to determine dangerousness? He has just convincingly explained how hard this is to predict and he offers no answer to that crucial question. Indeed, if dangerous behavior cannot be predicted with any confidence, how then could it be measured “to the inch”? How will future dangerousness be evaluated, and who will bear the burden of proof? Legal rules are designed to specify which side in a legal dispute should bear the risks of uncertainty and possible error. Thus, the legal standard for criminal convictions typically requires proof at the very high “beyond a reasonable doubt” standard, precisely to protect the rights of defendants in the criminal law system. Why would

34. SAPOLSKY, *supra* note 3, at 242.

35. *See id.*

36. *Id.* at 242–43.

37. *Id.* at 340.

38. *Id.* at 347.

39. *Id.*

40. *Id.* at 349.

41. *Id.*

42. *Id.* at 351.

throwing that law out help the most vulnerable defendants, who are typically the ones convicted of violent street crimes?

Procedures based on predicting future dangerousness have recently come into wide use in a variety of criminal law contexts, including pretrial detention, sentencing, parole, and probation.⁴³ Predictive models are also now being used for “predictive policing,” which seeks to identify which individuals are most likely to commit future crimes.⁴⁴ Experts have been highly critical of the results.⁴⁵

Predictive techniques in criminal law analyze large data sets to produce algorithmic models predicting individuals’ future likelihood of causing harm.⁴⁶ What do the strongest predictive factors turn out to be? Key factors are “neighbourhood deprivation,” as measured by how many people receive welfare payments along with divorce rates, education level, crime, and average community income levels; male sex; unemployment; youth; alcohol or drug use disorder; and mental disorder.⁴⁷ This list sounds a lot like the factors Sapolsky laments as making people statistically more likely to commit crimes for reasons that are not their fault.⁴⁸ In other words, assessing future dangerousness produces the same results as convicting persons who commit crimes, except without according defendants traditional due process rights. Put otherwise, predictive analysis does not solve the potential human rights violations embedded in the social and economic deprivations that often underlie the factors that turn out to be most predictive of future crimes. Under predictive harm-avoidance schemes as under traditional criminal law processes, factors that are not individuals’ fault—and result from a lack of protection of human rights to education, adequate income, health

43. See Shima Baradaran & Frank L. McIntyre, *Predicting Violence*, 90 TEX. L. REV. 497, 497 (2012); see also Melissa Hamilton, *Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law*, 47 ARIZ. ST. L.J. 1, 5 (2015).

44. Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1113–14 (2017).

45. See, e.g., Megan T. Stevenson & Sandra G. Mayson, *Pretrial Detention and the Value of Liberty*, 108 VA. L. REV. 709, 712 (2022) (describing the results as “terrifying” in failing to adhere to the “central constraint on criminal punishment” of proof of prior harmful acts).

46. Ferguson, *supra* note 44, at 1113.

47. Christopher Slobogin, *Principles of Risk Assessment: Sentencing and Policing*, 15 OHIO ST. J. CRIM. L. 583, 584 (2018) (describing the popular Oxford Risk of Recidivism Tool, which has identified a group within which sixty percent committed a violent crime in the next two years).

48. As we will discuss further in Section II.C below, this list also overlaps with needs, such as access to an adequate income for oneself and one’s family, and access to adequate medical care for physical and mental health, which are currently enshrined as universal human rights in international law. See G.A. Res. 217 (III) A, Universal Declaration of Human Rights, at 23, 25 (Dec. 10, 1948) [hereinafter *UDHR 1948*].

care, employment, and other social resources essential for human flourishing—increase the statistical likelihood that individuals will cause future harm to others (though, as Sapolsky points out, many individuals defy these predictions).

Sapolsky criticizes the racial bias endemic in criminal law today. We fully agree that this is a damning indictment that should spur criminal justice reform. But here is the problem for Sapolsky's proposal: evidence-based risk assessment turns out to be no better. One expert describes future harm prediction models as "an explicit embrace of otherwise-condemned discrimination, sanitized by scientific language."⁴⁹ Moreover, the factors that predict future dangerousness sometimes run counter to mitigating factors. For example, algorithmic risk assessment turns up youthfulness as a strong predictor of future violent criminal behavior, but under a "blameworthiness" assessment, youth is a mitigating factor—in part precisely because neuroscience (and common-sense observation) shows that young people's brains have not fully matured in the areas related to impulse control.⁵⁰ The same is true for mental illness, substance abuse, and lack of education;⁵¹ these factors mitigate blameworthiness but also predict future violent crime.⁵²

In short, we do not find Sapolsky convincing when he suggests that "quarantining" persons based on evidence of future dangerousness would make the system fairer. Quarantine facilities would remain full of the less privileged persons, often Black and Brown, whose life circumstances predispose them, statistically speaking, to commit future harms. Bias, in all its many forms, would pervade assessment. Sapolsky has elsewhere written wonderfully about the brain's in-group versus out-group dynamics,⁵³ but he barely acknowledges these problems with his proposal.⁵⁴

49. Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 803 (2014); see also Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2218 (2019) ("In a racially stratified world, any method of prediction will project the inequalities of the past into the future.").

50. Megan T. Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 WASH. U. L. REV. 681, 685 (2018) (finding age inversely explains fifty-seven percent of the variation in violence recidivism risk); see also *Miller v. Alabama*, 567 U.S. 460, 471–72 (2012) (proposing an inverse relationship between youth and degree of culpability). Youth is also a mitigating factor in international human rights covenants approved since 1948. See *UDHR 1948*, *supra* note 48; White & Gonsalves, *supra* note 6, at 46.

51. This is itself a human rights violation. See *UDHR 1948*, *supra* note 48, at 26.

52. Stevenson & Slobogin, *supra* note 50, at 687–88, 706.

53. See SAPOLSKY, *supra* note 3, at 387–424.

54. See *id.* at 352 ("This raises the specter of creepy precrime apprehension (as well as the need to keep an eye on the biases of the folks predicting the future criminality)."). Sapolsky dismisses the creepiness concern simply by pointing out that public health

To be sure, in this book, as in his others, Sapolsky's deeply ethical and humanitarian bent shines through. What he is fundamentally concerned about is the egregious injustice of throwing people in prisons because of their bad luck in being born into disadvantageous circumstances or with brains that make it impossible to avoid committing a harmful act. But that would seem to us to go to the question of whether there should be *more* exceptions to criminal culpability in situations of poverty, childhood abuse, mental illness, addiction, compulsion, traumatic brain injury, and genetically based brain anomalies. In other words, neuroscience could push law to adopt expanded exceptions to culpability based on new evidence, as, indeed, a large body of neurolaw literature is investigating.⁵⁵ International human rights law likewise

officials already use quarantines and does not return to the implicit bias problem at all. See *id.* at 352–53.

55. A classic in this genre is Deborah Denno, *Neuroscience and the Personalization of Criminal Law*, 86 U. CHI. L. REV. 359, 359–60 (2019) (arguing, on the basis of a large empirical study, that neuroscience evidence in particular cases may increase overall fairness in the criminal justice system). Other examples include Michael S. Moore, *Addiction, Responsibility, and Neuroscience*, 2020 U. ILL. L. REV. 375, 466 (2020) (arguing that neuroscience can deepen explanations of addiction and change evaluation of behavior but that its potential to expand the category of those excused by addiction rests largely in the future); Andrea L. Glenn & Adrian Raine, *Neurocriminology: Implications for the Punishment, Prediction and Prevention of Criminal Behaviour*, 15 NAT. REV. NEUROSCIENCE 54, 58 (2014) (arguing that the United States should move toward tests used in Netherlands for assessing degree of criminal responsibility based in part on neuropsychological assessment, but also noting that neuroscience is not yet poised to make immediate major changes in criminal law); Susan A. Bandes, *The Promise and Pitfalls of Neuroscience for Criminal Law and Procedure*, 8 OHIO ST. J. CRIM. L. 119, 120 (2010) (noting that neuroscience has been used in many aspects of the criminal justice system and has future potential uses); Richard E. Redding, *The Brain-Disordered Defendant: Neuroscience and Legal Insanity in the Twenty-First Century*, 56 AM. U. L. REV. 51, 53, 118 (2006) (arguing that neuroscience can help show that a defendant has limited physiological ability to control behavior and that defendants should be permitted to present evidence that mitigates their culpability based on impaired impulse control); Emad H. Atiq, *How Folk Beliefs about Free Will Influence Sentencing: A New Target for the Neuro-Determinist Critics of Criminal Law*, 16 NEW CRIM. L. REV. 449, 452, 483 (2013) (arguing that “[g]iven the plausible assumption that we have a collective moral duty to maintain a decent quality of life for the least well-off, our failure to fulfill that duty, insofar as it perpetuates criminality, undermines our moral standing to punish criminals with under-privileged backgrounds” and exploring empirical literature on the link between criminal behavior and deprivation in life opportunities); Jeffrey L. Kirchmeier, *A Tear in the Eye of the Law: Mitigating Factors and the Progression Toward a Disease Theory of Criminal Justice*, 83 OR. L. REV. 631, 636–37, 715 (2004) (arguing that law now seeks to better understand the causes of crime, and that examination of mitigating factors will become more important as scientific understandings begin to have a broad impact on law). In citing these examples, we are not intending to agree with or endorse the arguments they make, but simply to show the wide variety of scholarship underway on how neuroscience might be used to reform criminal law doctrines. A persistent theme across this literature is that the ability of neuroscience to

demands that criminal law embrace a far more compassionate understanding of the limits of human responsibility.⁵⁶ But that is a *very* different proposal than imposing an unabashed social control model based on quarantining persons who fail predicted future dangerousness tests.⁵⁷

Sapolsky points out that over millennia many societies have become more compassionate and realistic about the limits of human responsibility. For example, “whispers of modernity” could be heard as persons who committed violent acts during seizures started to be acquitted.⁵⁸ He believes that in another 500 years this process will be still farther along.⁵⁹ We agree: criminal law should show far more understanding of the limits to human responsibility and probably will in another half millennium. Where we part ways is on the question whether *all* persons should be viewed as having *no* responsibility for actions that cause harm to others, so that responsibility drops out of legal analysis completely. That would seem to mean that a person with many life advantages and no apparent brain damage (think Bernie Madoff) who showed no propensity for future violence would avoid quarantine, while a destitute person with the factors related to past social deprivation that predict future dangerousness would be locked up for life. Human moral intuitions (whatever their sources) bridle at such outcomes and, we believe, (but of course cannot prove, any more than Sapolsky can prove the opposite proposition) will continue to do so very far into the future.

In summary, we agree completely with Sapolsky’s call for more humaneness and compassionate understanding of the factors that should come into play in assessing human responsibility in criminal law, but on our view those very considerations call for continuing to distinguish between more and less culpable actors rather than eliminating questions of culpability entirely, as Sapolsky advocates.

contribute to criminal law reform is at this point only at early stages and much more scientific discovery is necessary before neuroscience can have a major impact.

56. UDHR 1948, *supra* note 48.

57. Cf. MINORITY REPORT (20th Century Fox 2002) (presenting a dystopian science fiction treatment of a police department that apprehends killers prior to their committing bad acts). Brilliant minds of past eras have grappled with similar questions of free will versus algorithmic prediction of an unfixed future. See, e.g., PHILIP K. DICK, *The Minority Report*, in THE MINORITY REPORT: AND OTHER CLASSIC STORIES 71 (Citadel Press 2002) (1956); Isaac Asimov, *All the Troubles of the World*, in IF THIS GOES ON 88 (Charles Nuetzel ed., Book Co. of Am. 1964) (1958). Note that the 20th Century Fox film was adapted from the Philip K. Dick novella.

58. SAPOLSKY, *supra* note 3, at 310.

59. Robert M. Sapolsky communication to authors, Jan. 24, 2024 (on file with authors).

B. Mitchell's Free Agents

Mitchell is a professor of genetics and neuroscience at Trinity College in Dublin, Ireland. Like Sapolsky, he has brought the insights of his field to a popular audience through previous publications.⁶⁰ In *Free Agents*, Mitchell offers a neuroscience perspective very different from Sapolsky's, based in part on current best understandings in evolutionary biology and neuroscience. As Mitchell explains, current best thinking about how life began on earth is that this happened in thermal heat shafts at the bottom of the ocean, where molecules began combining and then differentiating from other molecules through membranes; these became the first cells.⁶¹ This creation of an envelope differentiating living organisms from their environment is an especially important insight Mitchell offers, to which we will return in Section II.C.

But first, to briefly sketch Mitchell's explanation of how multicellular organisms developed and what this means for understanding agency: over billions of years, more complex living organisms began to evolve. What set these organisms on a path to survival was their ability to adapt to their environments.⁶² Certain evolutionary developments helped too. The "invention" of nervous systems allowed additional layers of processing in perceiving and evaluating information.⁶³ The ability to move helped too, allowing organisms to assess environments from different perspectives and calibrate the inferences they are making and adjust them if necessary.⁶⁴ Even at this point, the basic elements of what Mitchell calls "agency" are in place: i.e., there are self-sustaining systems, insulated from the environment, that work to keep themselves alive by choosing among possible actions after integrating information about internal and external states and interpreting the meaning and value of that information.⁶⁵

In other words, as Mitchell puts it, *reasons* motivate all living organisms' acts. Organisms extract *meaning*, bringing to bear prior experience and expectations as well as their evolutionarily selected genetic heritage.⁶⁶ Expectations matter too.⁶⁷ Organisms develop *beliefs* about what they are perceiving, about very simple things (am I seeing a

60. See, e.g., KEVIN J. MITCHELL, INNATE: HOW THE WIRING OF OUR BRAINS SHAPES WHO WE ARE (2018) [hereinafter MITCHELL, INNATE].

61. MITCHELL, FREE AGENTS *supra* note 4, at 19.

62. See *id.* at 68–95.

63. *Id.* at 67.

64. *Id.* at 68–69.

65. *Id.* at 68.

66. *Id.* at 117–18.

67. *Id.* at 118.

line or a shape?) as about complicated things, as in human belief systems.⁶⁸

Organisms must decide among multiple attractive stimuli, having to suppress some ideas and go with others.⁶⁹ Neurons do this work, sending patterns of mutual inhibitions that lead to some neurons firing and others being suppressed.⁷⁰ Living organisms, in other words, have the capacity for self-control; they are able to adjust their behaviors by electing the objects of their attention and prioritizing certain options for action.

Humans are really no different from other living organisms, just more complex. They can create goals beyond immediate survival, plan over longer time frames, and inspect their own reasons, i.e., they can think about thinking or engage in what Mitchell terms “metacognitive scrutiny.”⁷¹ Humans have the capacity for great creativity and creative problem solving,⁷² a point Sapolsky recognizes too. What is different in Mitchell’s framing is that, although the basis for all these capacities is neural and molecular, it is more than prewired instinct. The emergence of associative learning and long-lasting memory lets organisms make decisions based on their *own* reasons.⁷³ We will have more to say about this important point below.

Note that Mitchell’s description of how living organisms function is all it takes to satisfy the definition of agency that law requires. It does not matter *why* individuals make decisions about what actions to take and then execute those actions based on reasons, it just matters that they have the capacity to do this. Put most simply, regardless of whether their deliberating, choosing, and acting have prior causes all the way down or not, living organisms are *agentic* under the definition we offered in Part I above.

Mitchell further describes how human development happens through socialization. Caregivers encourage certain habits, which grow to become habits of character; these typically are prosocial traits.⁷⁴ These include self-control, patience, prudence, temperance, foresight, and perseverance.⁷⁵ The culture one grows up in provides moral instruction and feedback from external sources, which give individuals a basis to

68. *Id.* at 115–16.

69. *Id.* at 124.

70. *Id.*

71. *Id.* at 68–69.

72. *Id.* at 191.

73. *Id.* at 94.

74. *See id.* at 239–40.

75. *Id.* at 240.

evaluate their behavior with reference to social norms.⁷⁶ Internal feedback occurs through emotions associated with evaluating outcomes, such as regret, satisfaction, disappointment, frustration, and pride. Those states become learning signals for future behavior, and anticipation of emotions figures into considering options for action.⁷⁷

Both Mitchell and Sapolsky are interested in how moral sentiments evolved. Sapolsky focuses on the human impulse to impose retribution on those who transgress a society's morals, a sentiment he wants to extinguish, along with the converse impulse to admire and praise persons seen to engage in prosocial behaviors. As he sees it, these people deserve no more honor than anyone else since all actions have prior causes that they, in fact, do not control.⁷⁸ But Mitchell notes that praising those who possess admired personality traits encourages socially cooperative behavior, just as heaping opprobrium and retribution on those who have transgressed community norms communicates social meanings about what the group discourages.⁷⁹ In short, even if people do not truly "deserve" praise or punishment, these practices have important prosocial functions.

Essentially, both authors agree that what organisms do is the product of all the interactions their ancestors, and they themselves, had with their environments, combined with feedback from natural selection, social environments, individual learning, and knowledge about causal relations in the world.⁸⁰ What they disagree about is the import of these scientific understandings. And we want to add still a third way of understanding what neuroscience shows about human agency, drawn from a large literature specifically studying the phenomenon of agency at the neural level.

C. Spatial and Temporal Resolution of Human Agency: The Agentic Envelope

Both Sapolsky, in his attention to his assumed deterministic universe, and Mitchell, in his full-throated articulation of biological

76. *Id.* at 243.

77. *Id.*

78. See SAPOLSKY, *supra* note 3, at 126, 402.

79. MITCHELL, *supra* note 4, at 282–84. Unlike Sapolsky, Mitchell is not centrally interested in or outraged about how criminal law operates today (perhaps in part because he does not live in the United States). He has some to say on the subject, concluding that, basically, the principles of criminal law should stay in place, but that "extreme pathologies" should be a basis for criminal law excuse. *Id.*

80. *Id.* at 2–3.

agency, fail to articulate several key features of the dance between causes and choices in humans. We articulate these in four themes.

First, while Sapolsky's determinist universe would appear at first glance to be incompatible with human agency, we propose that he is simply using the wrong yardstick—or scale—for most meaningfully framing the question. The real question is actually: at what spatial and temporal scale does human agency occur? To be sure, when viewed from the physics of the cosmos—galaxies traveling a million miles per hour,⁸¹ planets circling, stars birthing, black holes rippling,⁸² and the universe ultimately collapsing upon itself or diluting to a quiet and icy end⁸³—human actions matter not one wit. Human action has no impact on planetary motion, the fate of the sun, or the motion of the stars. This is fairly indisputable, but it is also irrelevant to understanding human agency. At the level of the cosmos, our actions have no discernable impact, but we are looking for effects at the wrong spatial and temporal scale. Similarly, within the quantum realm, at the smallest spatial and temporal resolution of the cosmos, where matter is simultaneously waves and particles, probabilities are expressed as clouds, and the observed and the unobserved are in an endless, enigmatic dance beyond our understanding,⁸⁴ human action is without impact, again due to the mismatch in spatial and temporal scales.

The physicist Steven Hawking makes this point in his essay titled, *Is Everything Determined*.⁸⁵ There he discusses the implications of the uncertainty principle for “free will,” the predictability of human behavior, and law. Hawking argues that “[t]he concept of free will belongs to a different arena from that of fundamental laws of science. If one tries to deduce human behavior from the laws of science, one gets caught in the logical paradox of self-referencing systems.”⁸⁶ Hawking offers the important insight that “natural selection seems to lead to us adopting the effective theory of free will.”⁸⁷ Note how this dovetails with Mitchell's emphasis, as we discussed in Section II.B. Hawking further counsels that

81. The Milky Way travels at approximately 1.3 million mph. Andrew Fraknoi, *How Fast Are You Moving When You Are Sitting Still?*, ASTRONOMICAL SOC'Y PAC. Spring 2007, at 1, 5.

82. See Tony Phillips, *Black Holes: Feeling the Ripples*, NASA (Jan. 1, 2002), <https://dataverse.jpl.nasa.gov/file.xhtml?fileId=7355&version=1.1>.

83. See *What Is the Ultimate Fate of the Universe?*, NASA, https://wmap.gsfc.nasa.gov/universe/uni_fate.html (Feb. 20, 2024) (detailing that current bets are on ice rather than fire).

84. See generally STEPHEN HAWKING, BLACK HOLES AND BABY UNIVERSES AND OTHER ESSAYS, 909–14 (2011) (providing a general discussion of quantum theory).

85. *Id.* at 908–14.

86. *Id.* at 912.

87. *Id.* at 913.

“[o]ne has to keep the investigation of the fundamental laws of science and the study of human behavior in separate compartments. One cannot use the fundamental laws to deduce human behavior.”⁸⁸ In other words, in moving across knowledge disciplines, one runs the risk of assuming one can apply tools that are central to one field to another field when they may in fact be inapposite. This is what happens, we have just argued, in thinking about causation in terms of physics and the movement of objects in the universe. Those inquiries present a vastly different temporal and spatial scale and call for paradigms very different from thinking about human decision-making at the scale of right here and right now.

Neither Sapolsky nor Mitchell fully appreciate the import of studying human agency at the temporal and spatial scale of human life. Sapolsky leaves this point completely unsaid and Mitchell leaves it largely unexplored. The spatial and temporal scale in which agency can be properly understood is the scale in which human action actually occurs—*right here, right now, by us*. We propose that this is the spatial and temporal resolution at which the neurobiological engines of agency reside. While Mitchell outlines a path to this framing, which he couches within his discontent with determinism,⁸⁹ he does not give a name to this path for understanding agency. It is a focus on the scale of right here, right now, in individual lives, that allows evaluation of human choices as agentic.

We term this frame the “agentic envelope,” by which we mean behavioral choices individuals have control over in the here and now. The causal concerns Sapolsky raises against human agency—including genetics, which operate over millennia to generations to years; hormones, which operate over years to months to days; and physics, which operate over nanoseconds (quantal) to billions of years (cosmos)—lie outside this causal envelope.⁹⁰ Here we draw from Mitchell’s critical insight that, from its earliest inception as a single-celled organism, life has been *causally encapsulated*.⁹¹ Encapsulation of the biological machinery of life within a lipid membrane created something new: “autonomous entities, causally sheltered from the thermodynamic storm outside.”⁹² This causal sheltering allows the organism to be separate from its environment, establishing its ability to survive and thrive in its time and in that of its descendants, through the iterative vehicle of actions made for reasons.⁹³

88. *Id.*

89. MITCHELL, *supra* note 4, at x.

90. *See supra* note 25 and accompanying text.

91. MITCHELL, *supra* note 4, at xi.

92. *Id.* at 19.

93. *Id.*

This causal encapsulation, a key insight of Mitchell's book, has significant and direct application to understanding human agency, its scale and impact, and its rightful implications for civil, criminal, and human rights law. The concept of the agentic envelope, in which individual actions in the here and now have impact and for which individuals bear responsibility, is very much in keeping with conceptions of law, as we will discuss in Part III below. The agentic envelope provides a way of framing questions of human agency and rights and responsibilities.⁹⁴

The agentic envelope is also anchored in current empirically based knowledge about the neurobiology, from cells to circuits, of agency in humans. Some of this knowledge comes especially from three areas: studies of the role of autonomy in (1) learning, (2) action, and (3) emotion and motivation.⁹⁵ As to learning, the evidence documents the importance of autonomy to optimizing learning behavior and outcomes.⁹⁶ We see this throughout Mitchell's emphasis on actions organisms select for reasons. The role of autonomy in learning has been well established,⁹⁷ and Mitchell describes its role in survival, which he subsumes with reproduction under the term "persistence."⁹⁸ Autonomy is central to the survival of all organisms, from simple single-celled ones to much more complex multi-celled creatures.⁹⁹

A second literature on agency focuses on the neurobiology of volitional action, a theme Mitchell articulates well. Organisms (including but not limited to humans) are not input-output machines but are instead engaged in a process of meaning-building to permit specific actions in

94. Note we define the agentic envelope as *behavioral choices individuals have control over in the here and now*. Sometimes factors with broader time frames are deterministically causal, such as genetics or disabling brain damage, in which case responsibility should not be imposed. But sometimes proximate causation comes from an individual deciding with reasons, right here and right now, which course of action to take. It is that kind of agency we use the concept of the agentic envelope to refer to here.

95. See MITCHELL, *supra* note 4, at 1–23.

96. See *id.* at 21.

97. See White et al., *supra* note 6, at 5–6.

98. For definitions of persistence, see MITCHELL, *supra* note 4, at 19–20. Mitchell also describes the "self" in terms of persistence. See *id.* at 21 ("There is no self in a given moment: the self is defined by persistence over time.").

99. See *id.* at 52 ("Even better would be if the organism would have the ability to decide *where* to go. In an environment where resources and threats are unevenly distributed, it would pay to move toward the former and away from the latter, rather than just drifting aimlessly around. Such tendencies would certainly increase survival and thereby be selected for."). Note the similarity here to Hawking's point about the evolutionary benefits of believing in human agency; organisms that are active and strive to move in rational directions are more likely to survive. See HAWKING, *supra* note 84, at 70.

response to specific reasons.¹⁰⁰ For instance, in the realm of sensory perception, neural systems are not simply engaged in “information processing” for the sake of iterative elaboration per se. Rather, perceptual systems in even the simplest organisms tune to the creation of *meaning* with reference to the organism’s goals, varying from persistence (in amoeba, for example), to the much more elaborate goals of humans.¹⁰¹ Moreover, such meaning is not “extracted,” but rather constructed by the organism in light of its goals.¹⁰²

This perspective aligns with a third body of empirical literature on the contributions of “framing” to the interpretation and creation of emotion.¹⁰³ Humans’ individually specific creation of meaning provides the basis by which “talk therapies” work, for example, where cognitive reframing plays a substantive role.¹⁰⁴ This meaning-creation provides the basis for healing of clinical anxiety or negative emotion, for example, and also shapes more mundane, everyday emotional phenomena associated with the creation of meanings through other-regarding emotions such as compassion, empathy, and forgiveness.¹⁰⁵

Related empirical findings about human agency flow from the study of “agentic extraversion,” a construct drawn from human personality studies that speaks to individuals’ capacity for incentive motivation and

100. As Mitchell puts it, nervous systems extract meaning. See MITCHELL, *supra* note 4, at 22 (“[E]ven though our cognitive systems have a physical instantiation, their workings cannot be *reduced* to this level. We are not a collection of *mere mechanisms*. . . . [T]he nervous system runs on meaning.”).

101. On the relationship between goals and meaning, see *id.* at 43 (“[U]nlike the designed machines and gadgets that surround us in our daily lives, which also have a purpose or at least serve a purpose, living organisms are adapted for the sake of only one thing – their *selves*. This brings something new to the universe: a frame of reference, a subject. The existence of a goal imbues things with properties that previously never existed *relative to that goal*: function, meaning, and value.”).

102. On the difference between information and meaning in biological systems, see *id.* at 63–64 (“[I]t is meaningful information—information *about things*—that living organisms need when trying to infer what is out in the world.”).

103. Creation of meaning is also important in conscious experience, such as subjective emotion in response to a given event, which differs across people as a function of past experience. See LISA FELDMAN BARRETT, *HOW EMOTIONS ARE MADE* 25–41 (2017) (discussing such “constructivist” theories of emotion).

104. *Id.* at 211. For an overview of cognitive behavioral therapy, see Gianluca Serafini et al., *Overall Goal of Cognitive-Behavioral Therapy in Major Psychiatric Disorders and Suicidality: A Narrative Review*, 107 MED. CLINICS N. AM. 143, 143–44 (2023).

105. See BARRETT, *supra* note 103, at 246; Tania Singer & Olga M. Klimecki, *Empathy and Compassion*, 24 CURRENT BIOLOGY R875, R875–76 (2014); see also Melike M. Fourie et al., *Parsing the Components of Forgiveness: Psychological and Neural Mechanisms*, 112 NEUROSCIENCE BIOBEHAVIORAL REV. 437, 437–39, 443, 448 (2020).

emotional states that foster action.¹⁰⁶ This literature also studies differences in its expression among various persons.¹⁰⁷ Agentic extraversion represents abilities to experience feelings of self-efficacy, work toward goals, recruit resources (often including the motivation of other people), and extract oneself from negative circumstances, such as danger and abuse.¹⁰⁸ Similar to Mitchell's emphasis on life as constructing meaning, having goals, and acting for reasons, agentic extraversion is the expressed predisposition to engage in the activities just listed, and its study aims to identify what internal, subjective states facilitate these processes. These states can have both positive and negative valence, including positive emotions such as elation, exuberance, and persistence, as well as aversive states such as anger, frustration, and determination in the face of adversity.¹⁰⁹ All of these emotional states facilitate forward action toward goals, including difficult or perilous ones.¹¹⁰ In short, the basis for human agency has been empirically studied through investigations of the neurobiology of autonomy in learning; of agency, meaning and value in transitions from perception to motivation and action; and of positive and negative emotion and motivation on display in the personality trait of agentic extraversion.¹¹¹

A large scientific literature further indicates that persons who believe they personally have no agency in their lives experience profoundly negative effects on their well-being. This literature is different from the studies of students who are read passages refuting belief in free will that Sapolsky discusses,¹¹² and to our minds the literature we mention here is of far more consequence in studying persons' actual life experiences rather than a few moments in an artificially contrived experimental setting. The negative effects of

106. See Richard A. Depue & Paul F. Collins, *Neurobiology of the Structure of Personality: Dopamine, Facilitation of Incentive Motivation, and Extraversion*, 22 BEHAV. BRAIN SCI. 491, 518–55 (1999) (discussing agentic extraversion).

107. See *id.*

108. See *id.* at 495–55.

109. See Jeannine V. Morrone et al., *Film-Induced Incentive Motivation and Positive Activation in Relation to Agentic and Affiliative Components of Extraversion*, 29 PERSONALITY & INDIVIDUAL DIFFERENCES 199, 200 (2000) (discussing positive emotion); see also White et al., *supra* note 6, at 2 (discussing negative emotion).

110. See White et al., *supra* note 6, at 2. Phenotypic differences in self-control demonstrate the biological basis of agency. See MITCHELL, *supra* note 4, at 272 (“[D]iverse types of individual differences provide another source of evidence that what we think of as free will – the ability to consciously monitor and control our cognitive processes, to reason about our reasons, and thus regulate our own behavior – is not some abstract metaphysical postulate but an evolved function, or suite of functions, with a very real biological basis.”).

111. See White & Gonsalves, *supra* note 6, at 41–44.

112. SAPOLSKY, *supra* note 3, at 243–44.

individuals' beliefs that they are helpless, and thus hopeless, defined as belief in one's inability to impact events,¹¹³ are well documented; reduction in beliefs about one's helplessness improves well-being.¹¹⁴ Even where agency appears to be truncated or short-circuited, such as in substance dependence, the presence of the "will to will," termed "second-order desire"¹¹⁵ or "second-order volition" for ceasing an addictive practice, is an important factor in moving out of addiction.¹¹⁶ Treatment methodologies work to provide systems and procedures to support this "second will"; popular twelve-step communities, such as Alcoholics Anonymous and Narcotics Anonymous, rely upon personal connections and community to actively support the "will to will"—here, the drive to maintain sobriety in the face of cues that would otherwise overwhelm or override agency.¹¹⁷ Efforts to predict and prevent relapse are similarly aimed at supporting and honoring this "second level" of will.¹¹⁸

Still other literatures investigate agency or empowerment in the oppressed, marginalized, and abused, such as in human rights for refugees, protection of minorities from discrimination, and empowering women to exit abusive environments.¹¹⁹ The view that human agency is an illusion suggests that change is not possible, and outcomes cannot be altered; thus, why should anyone try? This view eviscerates both

113. These terms have distinct yet overlapping use in the literature. See Steven F. Maier & Martin E. P. Seligman, *Learned Helplessness at Fifty: Insights from Neuroscience*, 123 PSYCH. REV. 349, 349 (2016) (discussing helplessness); Richard T. Liu et al., *The Hopelessness Theory of Depression: A Quarter-Century in Review*, 22 CLINICAL PSYCH.: SCIENCE & PRACTICE 345, 246 (2015) (discussing hopelessness).

114. See Melanie A. Price et al., *Helplessness/Hopelessness, Minimization and Optimism Predict Survival in Women with Invasive Ovarian Cancer*, 24 SUPPORT CARE CANCER 2627, 2627 (2016) (discussing the preventative impact of optimism on mortality in cancer).

115. See McKenna & Coates, *supra* note 14, at 28–29. As these authors explain, second-order desires are desires for effective first-order desires that would comprise "will," and thereby be effective in moving an individual all the way to action. "For instance, the dieter who is constantly frustrated by their sugar cravings might desire a more effective desire for health, one that would be more effective in guiding their eating habits than it often is." *Id.* at 28.

116. See Gene M. Heyman, *Addiction and Choice: Theory and New Data*, FRONTIERS IN PSYCH., May 2013, at 1, 4 ("[C]hoice models . . . predict that individuals caught in a destructive pattern of behavior retain the capacity to improve their lot and that they will do so as a function of changes in their options and/or how they frame their choices.").

117. John F. Kelly et al., *Alcoholics Anonymous and Other 12-Step Programs for Alcohol Use Disorder*, COCHRANE DATABASE SYST. REV. no. 3, 2020, at 1, 15 (explaining how twelve-step programs "increase psychological well-being, improve interpersonal skills, enhance the ability to cope with stress, and facilitate adaptation to abstinence and a sober lifestyle").

118. See, e.g., *Projects for the Development of Safety and Healthcare: Addiction Treatment Program*, STRATEGIC AID PARTNERS, <https://strategicaidpartners.com/projects/#Addiction> (last visited Feb. 16, 2024).

119. See generally White et al., *supra* note 6.

meaning and hope for the future. It also contradicts what the empirical evidence shows about the impact of rights-based freedoms in a healthy democracy.¹²⁰ Agency and autonomy are also important themes in the emerging neurorights movement,¹²¹ which calls for rights to cognitive liberty,¹²² autonomy-based rights, and brain health protections as policies supporting a flourishing society.¹²³

To be useful, a concept must be well-defined. Such definition typically follows an iterative course, with initial constructions leading to more effective, and more accurate, delineations. This process requires a recursive process of reflection, input and—for science—iterative empirical testing. Thus, it is with no small amount of humility and openness to risk that we explore the nature of human agency and its potential applications to law and other human endeavors. The concept of the agentic envelope draws on the insights of Mitchell in his exploration of the causal encapsulation of biological life from its immediate environment, beginning with single-celled life and leading to more complex organisms, such as humans. This conceptualization, narrowly construed, would suggest that the agentic envelope ends where the other begins—i.e., at the boundaries of the confines of our individual human bodies. This definition would be consistent with legal rules that hold individuals responsible as individuals. With this narrow construction, the agentic envelope would be consistent with existing practice and theory in law.

However, unlike single celled organisms, in humans the agentic envelope is permeable and connected to the surrounding social context. As both Mitchell and Sapolsky point out, each of us hold within ourselves—within the discrete human package we refer to as the agentic envelope—the multitude of our experiences and the universe of our

120. See Sebastian F. Winter et al., *Brain Health-Directed Policymaking: A New Concept to Strengthen Democracy* 2 (Brookings Inst., Working Paper No. 178, 2022), https://www.brookings.edu/wp-content/uploads/2022/11/Brain-health-directed-policymaking_Final.pdf (outlining how “brain health” policy can support a vibrant economy and democracy and how brain health challenges, including mental, neurologic, and substance use disorders, and social determinants of healthcare associated with “substantial economic and sociopolitical impediments”).

121. For a review of the history of neurorights and its relation to neurolaw, see Marcello Ienca, *On Neurorights*, FRONTIERS HUM. NEUROSCIENCE, Sept. 2021, at 1; see also Nita Farahany, *Cultivating Cognitive Liberty in the Age of Generative AI*, in AI ANTHOLOGY (Eric Horvitz ed., 2023), <https://unlocked.microsoft.com/ai-anthology/nita-farahany>.

122. Farahany, *supra* note 121.

123. See Winter et al., *supra* note 120, at 2 (“[T]hriving democracies can distinguish themselves through provision of environments that enable each citizen to achieve their full brain health potential conducive to both personal and societal well-being. Gearing policymaking towards equitable and quality brain health may prove essential to combat brain challenges, promote societal cohesion, and boost economic productivity.”).

hopes.¹²⁴ These experiences and hopes live within the brain and body, but they are also the way in which the environment—throughout our development—shapes and changes us. In this way our past experiences, ideas discussed with others, and received wisdom (or lack thereof) all funnel into our agentic envelope to become part of the mix of choices and actions that create the possibilities for our future. Here a visual is useful—the “island of knowledge”: the more you learn, the bigger the island of knowledge gets—as so does the coast, that ever-widening place where our knowledge touches the unknown; education and training enlarge the island, but also enlarge the coast, resulting in the humbling (and accurate) realization of the magnitude of what one does not yet know; the coastline touching the ocean is like the agentic envelope and our impact on the world; education, experiences, training, and community all expand the island, shaping and increasing personal opportunities for action.¹²⁵ In this way, experiences are represented in brain and body, and affect the expanse and impact of the human agentic envelope.

And what about outputs? What is the impact of agency in a human life? As Mitchell and Sapolsky again discuss, humans are equipped with (and evolved for) collective action.¹²⁶ We see this in the mirror neurons in the brain,¹²⁷ the importance of social grooming and C tactile fiber somatosensory tracts for social cohesion,¹²⁸ our ability to feel—and

124. MITCHELL, *Free Agents*, *supra* note 4, at 21; SAPOLSKY, *supra* note 3, at 3.

125. See RALPH WASHINGTON SOCKMAN, *THE DOG AND THE MANGER* 202 (1946) (“[T]he field of knowledge which even the best of us can master is like an island surrounded by a limitless ocean of mystery. And the larger the island of knowledge, the longer the shore line of wonder.”). The concept of islands of knowledge is also expressed in work by the physicist and natural philosopher Marcelo Gleiser. See MARCELO GLEISER, *THE ISLAND OF KNOWLEDGE: THE LIMITS OF SCIENCE AND THE SEARCH FOR MEANING* xxii (2014) (“[A]s the Island of Knowledge grows, so do the shores of our ignorance—the boundary between the known and the unknown.”).

126. MITCHELL, *Free Agents*, *supra* note 4, at 19–20; SAPOLSKY, *supra* note 3, at 3.

127. Luca Bonini et al., *Mirror Neurons 30 Years Later: Implications and Applications*, 26 *TRENDS IN COGNITIVE SCI.* 767, 767 (2022) (“The mirror mechanism allows a basic and evolutionary widespread remapping of other-related information onto primarily self-related brain structures, in a large variety of domains, with a major role in social cognition and in guiding social interactions.”).

128. Laura C. Grandi, *From Sweeping to the Caress: Similarities and Discrepancies between Human and Non-Human Primates’ Pleasant Touch*, 7 *FRONTIERS PSYCH.* 1, 1 (2016) (“The static touch responsible for the discriminative aspect activates the large myelinated low threshold mechanoreceptors (LTMRs) to allow the rapid encoding of an object’s features at the central nervous system level. Conversely, the affiliative touch activates the C tactile unmyelinated LTMRs (CT fibers), to instead allow the processing of the emotional meaning of the touch.”).

learn—empathy,¹²⁹ and the capacity to act on behalf of others as well as ourselves (see literature on moral anger).¹³⁰ Our ability to leverage action depends, in part, on the strength, breadth, and depth of our relationships with others; illustrated by, for instance, the engagement of an influencer's followers; the effectiveness of a businessperson's network; the enthusiasm of a pastor's flock; the dedication of a crime boss's organizational network. In this way, the agentic envelope has both inputs (experience, knowledge, history, resources, community) and outputs (reach, engagement, networks) through which individuals, as causal agents, impact the world.

The concept of the agentic envelope provides a new way of thinking about the scope of action and effects, in which human experiences and leverage cohere in real time to shape the subject's and others' futures. On the "output" side, when an individual reaches out to others to join forces in action, their agentic envelope not only expands but transforms into a force multiplier. In the area of social justice, the concept of the agentic envelope provides hope and spurs action. In the words of John Lewis, "[g]et in good trouble, necessary trouble, and redeem the soul of America."¹³¹

In short, the agentic envelope provides a way to support human agency and dedication to action—whether big or small—in a world that is constantly finding new ways to disenfranchise individuals and tell them their actions do not matter. Without proof of this proposition, which science cannot provide, as we have already discussed, the metaphysical stance that there is no room for human agency appears to us to have significant potential to devalue and dehumanize human beings; this is the reason the role of agency is such an important aspect of concepts of intrinsic human dignity and universal human rights.¹³² As Mitchell puts it: "Agents themselves can be causes."¹³³

129. Cecilia Heyes, *Empathy Is Not in Our Genes*, 95 NEUROSCIENCE & BIOBEHAVIORAL REVS. 499, 499 (2018) ("I present a dual system model that distinguishes Empathy, an automatic process that catches the feelings of others, from Empathy, controlled processes that interpret those feelings. Research with animals, infants, adults and robots suggests that the mechanism of Empathy, emotional contagion, is constructed in the course of development through social interaction. Learned Matching implies that empathy is both agile and fragile. It can be enhanced and redirected by novel experience, and broken by social change." (internal citations omitted)).

130. See White et al., *supra* note 6, at 2, 17.

131. Rashawn Ray, *Five Things John Lewis Taught Us About Getting in "Good Trouble"*, BROOKINGS INST. (July 23, 2020), <https://www.brookings.edu/articles/five-things-john-lewis-taught-us-about-getting-in-good-trouble/>.

132. See White & Gonsalves, *supra* note 6, at 50.

133. MITCHELL, *Free Agents*, *supra* note 4, at 281.

III. LAW

What remains now is to bring law into the conversation with the several neuroscience perspectives discussed above. To do so, we offer simplified explanations of law, just as Sapolsky and Mitchell do for science. As we explain, law uses the concept of human agency to support the linked legal concepts of individual *rights* and *responsibilities*. Government supports individuals' rights to decide for themselves what they want to do, and individuals are held responsible for their decisions and actions if they cause harm to others.

A. *The Basics of Corrective Justice*

At the most basic level, law starts with the assumption that those who commit acts that harm others should compensate their victims for the harm caused. Why? Because someone has to bear the cost of the harm caused, and making the person who committed it pay corrects the harm.¹³⁴ This is the basic idea embodied in the term “corrective justice.”¹³⁵ There are also arguments that making people pay for the harms they cause incentivizes them to avoid many harms, so they do not have to pay for them. One potential approach is called “strict liability,” because it does not provide any excuses; essentially, if you did it, you pay.¹³⁶ Alternatives include an insurance system or taxes, where everyone shares the cost of compensating for harms. Another option is letting losses fall where they occur—essentially telling people who experience a harm caused by another, “tough luck.” But those alternatives either do not correct for harms or do not incentivize actors to avoid harm.

Another alternative introduces the concept of *fault*, thus introducing the notion of blame. It is important to recognize that this rule *favors* those whose actions result in harm to another, by *limiting* when one is liable to pay the costs of harms. Outside of criminal law, fault is defined in terms of negligence: the general rule of negligence in tort law holds that a person is at fault and thus liable for a harm they cause if a reasonable (generally defined as “average”) person in the defendant's situation would have taken more care to prevent the harm.¹³⁷ Note that fault comes

134. See RICHARD A. EPSTEIN, TORTS 87 (1999) (“[A]s between the two parties . . . it is better that the liability be imposed on the party who caused the harm rather than on the innocent victim who suffered it.” (internal quotations omitted)).

135. *Id.* at 86–87.

136. *Id.* at 69–70.

137. *Id.* at 110–12 (showing that law recognizes that for persons who fall below average in their abilities to avoid harming another, this rule continues to impose strict liability).

in to limit the circumstances in which a person whose acts cause harm will be held responsible.¹³⁸ This describes, in very simple terms, the civil law system for compensation for harms.

Criminal law, which deals with harms to society itself, has bigger teeth than civil law: it has the power not only to require restitution but also to take away individuals' fundamental right to freedom (about which we will say more below).¹³⁹ For this reason, standards for liability and proof in criminal cases are much *higher* (i.e., more pro-defendant) than standards for liability and proof in civil law. Prosecutors must establish beyond a reasonable doubt (as opposed to the "more likely than not" standard in civil cases), not only that defendants committed the acts for which they are charged, but also that they did so with what is called mens rea. Mens rea is a complex concept and the level of mens rea required varies crime by crime. Most simply put, for many crimes, mens rea boils down to intending, with bad motives, to bring about the harm they commit.¹⁴⁰ For many crimes, a defendant must have been either reckless or have committed the act with the specific intent to bring about the harm that occurred.¹⁴¹ Again, people are held responsible for the acts they commit, *but under more limited circumstances*. This is responsibility at its most basic level in criminal law. Note that establishing free will is in no way involved.¹⁴² Put otherwise, criminal law's frequent references to free will refer to something other than the metaphysical debate; it would be preposterous for a court to ask a party to "prove up" the free will/determinism proposition.

Where free will comes up is in the *defenses* to criminal liability. One defense is duress: if someone commits a criminal act because someone else is holding a gun to their head, then extenuating circumstances excuse their bad act.¹⁴³ But duress is a very limited defense; extreme poverty, or what the law calls economic necessity, has been held not to be a legitimate basis for a claim of duress, even though the need to eat in order to survive does, for many, seem to be a pretty good defense to the

138. *Id.* at 94. Thus, a prominent U.S. legal historian has pointed out that negligence rules, which became popular along with the growth of large industry, subsidize big businesses by partially immunizing them from having to compensate for harms their activities cause to workers and others. See MORTON HOROWITZ, THE TRANSFORMATION OF AMERICAN LAW, 1780–1860, 99 (1977).

139. JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 16–22 (8th ed. 2018).

140. *Id.* at 113–30.

141. *Id.* at 123–25.

142. *Id.*

143. *Id.* at 283.

act of stealing a loaf of bread.¹⁴⁴ Another defense is insanity.¹⁴⁵ Various jurisdictions have adopted varying formulations of the insanity defense, some broader and some narrower. These have fluctuated over time, influenced by the politics of those making the rules (mostly state legislatures, not always bastions of progressivism). Typically, succeeding on this defense does not lead to a better outcome for a defendant, however. Instead, the case moves into the involuntary civil commitment system, where persons who have committed harmful acts are often confined indefinitely with fewer due process protections than those accorded in the criminal law system.¹⁴⁶ A lesson for Sapolsky here: be careful what you wish for because alternatives to criminal law may be even worse.

B. Rights and Responsibilities as a Package Deal

Criminal law imposes responsibilities on individuals to not commit acts that legislatures elected by the people have defined as harmful and thus illegal. But *why* should persons have such responsibilities to avoid committing acts thus defined as harmful, aside from the assumption that punishing persons who engage in harmful acts tends to deter those who have the capacity to act on the basis of reasons from committing them? Part of the answer is: because they also have the right to make decisions for themselves.

This story may best be told by starting with John Locke, whose political theory is hugely influential in Anglo-American law. Like other philosophers of his time, Locke engaged with metaphysical arguments about free will, but he argued that the concept was “unintelligible,” and that the matters that really needed addressing were freedom and liberty, which, he said, “refer to the agent and not the ‘will.’”¹⁴⁷ Liberty, Locke proposed, “consist[s] in a power to act or to forbear acting, and in that only.”¹⁴⁸ Whether one is truly “free” in making a choice to act did not matter to Locke; the ability to engage in deliberative thinking was all Locke thought the term free will needed to mean. Locke also emphasized the ability to postpone decisions long enough to reflect on the

144. See Meir Dan-Cohen, *Responsibility and the Boundaries of the Self*, 105 HARV. L. REV. 959, 997–98 (1992).

145. DRESSLER, *supra* note 139, at 335.

146. *Id.* at 333–35, 348.

147. JOHN LOCKE, COMPLETE ESSAY ON HUMAN UNDERSTANDING, Ch. XXI, §§ 22, 24 (1836).

148. *Id.* § 24.

consequences of a choice.¹⁴⁹ Note the general convergence between Locke's deliberation and Mitchell's agency in living organisms.

What is especially important about Locke's views is how they interrelate with his political philosophy emphasis on the civil liberty to pursue one's self-defined goals (i.e., to act within one's agentic envelope) without undue government interference, provided that in doing so one does not unduly interfere with others' rights to do the same.¹⁵⁰ This was an important theme undergirding the U.S. founders' construction of government: individuals are better able to decide how they should live their lives than is government.¹⁵¹ To be sure, individuals should foster the virtues of character that help them make wise decisions,¹⁵² but once they are adults, they should make decisions for themselves. Reasonable laws can limit individuals' liberty as necessary to protect the public interest and other individuals' safety and security, but that's all.¹⁵³ In short: limited government plus free actors (operating within reasonable constraints prohibiting harm to others' rights) equals a legitimate, more or less "optimally" balanced, political and legal system, in which constituent actors make their own action decisions within reasonable constraints aimed at protecting others' rights to do the same.¹⁵⁴

If Sapolsky thought more about all this, we believe he would want this system. We can readily imagine him pointing out that, in reality, this system exists only for those from socially privileged backgrounds; we agree wholeheartedly with this point. But the appropriate analytical response to it is to insist on improving rights and opportunities rather than throwing out a system designed to protect rights (even though it fails to adequately do so) in favor of policing predicted future dangerousness. Here is another way of putting this point: under Sapolsky's social control system, under which no one "deserves" any rights, there are no ideas standing against the government determining all persons' futures, whether those futures involve being "quarantined" based on factors predicting future dangerousness or being assigned any other destiny, such as pursuing science with the goal of helping humanity. Remove the responsibility to be held responsible for one's acts

149. *Id.* § 66.

150. *Id.* § 24.

151. For discussion of the founders' conceptions of human agency, see Carle, *Founders' Conceptions*, *supra* note 7, at 542–50.

152. *See id.*

153. *See* 1 WILLIAM BLACKSTONE, COMMENTARIES 7 (Albany, Banks & Brothers, abr. 3rd ed., 1894) (1771).

154. *See supra* text accompanying note 25. (Note the consilience with the (more or less) optimal results achieved through the operation of individual acts in the emergent complex systems Sapolsky discusses).

and individuals' rights to self-determination goes out the window too. Rights and responsibilities are flip sides of the same coin. From this perspective, taking away the civil liberties that prevent the government from confining persons based on predictions of future dangerousness is a step toward less justice, not more.

C. Reading Neuroscience and Law Together

Scholars who bridge law and neuroscience have been making arguments adjacent to ours for some time. Among these is Professor Stephen Morse, a leading voice in neurolaw now mentoring a new generation. Morse has long argued that the free will/determinism debate does not defeat the basic assumptions of criminal law. As he has long argued, all that criminal law needs to impose personal responsibility is for persons to be able to act for "reasons."¹⁵⁵ The brain mechanisms that explain why they act for reasons, i.e., the *why* underlying their actions based on reasons, do not really matter; what matters is that individuals are "capable of acting for reasons and are capable of minimal rationality."¹⁵⁶ Hence the criminal law defenses for those who are not thus capable,¹⁵⁷ as we have briefly touched on in Section II.A above. But generally speaking, persons are "creatures who can act for and respond to reasons," and what law does is guide action by providing them with good "reasons for forbearance or action."¹⁵⁸

Morse points out that a significant part of the "misunderstanding and confusion" that occurs when members of different academic disciplines try to talk to each other about the concept of free will arises from the different understandings they have about this concept. In law, the term free will is synonymous with a person being criminally responsible.¹⁵⁹ It is a *different concept* than the libertarian free will philosophers' debate.¹⁶⁰ As Morse asserts, the claim that human beings "have the capacity to be guided by reason" is "fully consistent with the truth of determinism"; thus neuroscience "poses no challenge to our responsibility doctrines and practices."¹⁶¹

155. See Morse, *Free Will*, *supra* note 10, at 251.

156. *Id.* at 256.

157. *Id.* at 257.

158. *Id.* at 255.

159. *Id.* at 260. Morse discusses examples of how the mental conditions that can affect imposition of responsibility in criminal law stem from problems with rational capacity. See *id.* at 276, 278, 280.

160. *Id.* at 261.

161. *Id.*

Morse perhaps does not elaborate on what he means by individuals acting on the basis of “reasons” as extensively as one might like, but, as it turns out, Mitchell provides a scientifically grounded, biologically based elaboration of just that topic, explaining how living organisms assess, deliberate, choose courses of action based on reasons, and then execute decisions, as discussed in Section II.B above. In human affairs, law provides reasons for action and forbearance from action and provides limited defenses from liability for those who cannot act based on reasons. Grounded most basically in corrective justice, the law of personal responsibility reflects an inherited set of interlocking ideas that provides a cultural landscape for decision-making. That landscape requires analysis on its own terms. But we also want to suggest that the concept of the agentic envelope can do even better, as we discuss below.

D. Utility, Aperture, and Applications of the Agentic Envelope in Law

From a utility perspective, the agentic envelope may provide a helpful supplement to current understandings of agency in law. Its first strength is conceptual. The concept of an agentic envelope comes from knowledge based in the life and biological sciences. Just as Mitchell does in his treatment of agency generally, the concept of the agentic envelope knits together the concept of agency and processes scientists conceptualize in biology and physics. The agentic envelope provides a frame through which to reconcile known determinist processes in physics, from the atomic to the cosmic scale, to the processes humans encounter in the course of everyday life. In this way, the concept carves out a space in which we can understand and discuss how individuals take meaningful action within their own lives, despite the whirling chaos beyond.¹⁶²

Applications of the concept operationally can also shed light on phenomena that are otherwise difficult to reconcile. The breadth or narrowness of the aperture with which the agentic envelope is viewed is likely to correspond to the openness—indeed, the open-mindedness—of the individual standing in judgement of another’s actions, both within and beyond the field of law. Thus, applications of the agentic envelope may largely reflect the aperture of information—and the temporal scale—being used to evaluate actions, responsibility, and adjudicated consequences. The aperture of the information to be considered in assessing the extent of an individual’s agency relates both to the concept of the agentic envelope, in its temporal and spatial resolution, and to the

162. See MITCHELL, *Free Agents*, *supra* note 4, at 19 (referring to the “thermodynamic storm outside”).

decision-makers' and legal system's openness to understanding and investigating circumstances that may mitigate, explain, or provide extenuating circumstances relevant to judging actions in particular real-world contexts.

The aperture used to evaluate the agentic envelope will take into account, to a greater or lesser extent, the political landscape in which decisions and actions, motivation, and deserved consequences occur. The scope of this aperture may determine the legal consequences for an infraction. Here we consider a thought experiment: a survivor of domestic violence who kills their abusive spouse. A narrow aperture on the lens of spatial and temporal resolution of human action would lead to the judgment that "they killed their spouse, they must go to prison, end of story"; an open and shut case of immediate actions and consequences. A wider, more-opened aperture would consider the decades of abuse, the immediate threat of bodily harm, and prior and current actions taken in self-defense. Understanding the contributions of temporal and spatial specificity of agency—i.e., using and varying the temporal and spatial bounds of the agentic envelope—provides insight on the ways in which the aperture of judgment and resulting consequences plays out in response to specific human actions, times, and contexts. In this way, the agentic envelope can include political priors, lived experience, and cultural context. The concept of the agentic envelope argues against a foreshortened temporal and spatial frame for positioning the agentic envelope in law. Such a truncated frame would do great disservice to understanding and accounting for systemic inequities and assaults to intrinsic human dignity over time, both in the moment and in the periods leading up to—and flowing from—specific choices and actions. In this way, the concept of the agentic envelope, as further developed and construed, stands in service of bringing about legal reforms that would better promote justice and accountability in directions aligned with and responsive to the needs of contemporary understandings, and do so in order to better focus on the perspectives of those least able to access justice and remediation in the current system, a central problem for Sapolsky as well, as we have already noted.

The agentic envelope thus provides a useful aperture and lens through which to understand and consider the factors, legal and otherwise, that lead to actions in real-world contexts. We reconcile concepts of agency as embedded in real-world contexts through this sketch of the concept of the agentic envelope. This envelope provides a bridge between biology and law in understanding the emergence—and

reality—of agency in a universe full of causes,¹⁶³ with human agency itself emerging as a causal factor¹⁶⁴ at the spatial and temporal scale of human life.

CONCLUSION

We have argued that Sapolsky's exegesis of current thinking about chaos theory and complex emergent systems itself demonstrates the problems with his criminal justice proposal. Throwing out all defenses based on lack of capacity does not humanize the criminal justice system; it instead calls for imposing social control based on highly imperfect predictions of future dangerousness. What Sapolsky's explanations show, even while he argues the opposite, is that science cannot solve the free will versus determinism debate. Those big issues remain in the realm of metaphysics for at least the foreseeable future. They do not affect law's legitimacy in any case, since law is not interested in "free will" in the metaphysical sense (as shown, if by nothing else, by the fact that courts do not call on litigants to prove up this matter).

Mitchell, in turn, offers an explication that highlights a consilience between certain strains of neuroscience and law. The thesis of his book—that all living organisms act for reasons—supports a key point neurolaw scholars such as Morse have been making for some time, which is that all law needs for its legitimacy is that individuals act with reasons. *That* is what "free will" means in law. In other words, what law needs from the metaphysical free will concept is quite minimal, in the form of the proposition that human beings (and all living organisms) are *agentic*.

We have also offered a third way of thinking about agentic action, based on the concept of the *agentic envelope*. This perspective highlights

163. Sapolsky posits that most, if not all, arguments for "free will" require nondeterminism. SAPOLSKY, *supra* note 3, at 3. In a chapter entitled "Turtles all the way down," Sapolsky explains that:

[W]hen people claim that there are causeless causes of your behavior that they call "free will," they have (a) failed to recognize or not learned about the determinism lurking beneath the surface and/or (b) erroneously concluded that the rarefied aspects of the universe that do work indeterministically can explain your character, morals, and behavior.

Id. We disagree with this view and argue that human agency itself is one of the causes shaping everyday life. Further, understanding the "how" of any action, which Sapolsky characterizes as involving "turtles all the way down," does not subvert or even address the "why" of an action, for which the concept of agency provides both engine and cause. *Id.* at 1.

164. See MITCHELL, *Free Agents*, *supra* note 4, at 273 (explaining, in chapter entitled "Free Will," that "[t]he story of agency is really the story of life itself"); *see also id.* at 281 ("[H]igher-order organization can be part of the cause of things that happen. Agents themselves can be causes.").

strains of neuroscience that investigate the importance of agency for human flourishing and well-being. Together, these several varying approaches demonstrate the fruitfulness of engaging in dialogue across intellectual differences, not only across the disciplines of metaphysics, neuroscience, and law, but also within disciplines. Differences in vocabularies, assumptions and needs can lead to counterproductive confusion when participants make translation errors across different knowledge domains. At the same time, cross-pollination among discourse communities can contribute to advancing knowledge, and we greatly look forward to much more such interdisciplinary discourse on human agency, law, and personal responsibility in years to come.

