An Evidence-Based Objection to Retributive Justice

Brian T.M. Mammarella*

Abstract:

Advancements in neuroscience and related fields are beginning to show, with increasing clarity, that certain human behaviors stem from uncontrolled, mechanistic causes. These discoveries beg the question: If a given behavior results from some combination of biological predispositions, neurological circumstances, and environmental influences, is that action unwilled and therefore absolved of all attributions of credit, blame, and responsibility? A number of scholars in law and neuroscience who answer "yes" have considered how the absence of free will should impact criminal law's willingness to justify punishments on the basis of retribution, with some arguing that criminal law ought to dispense with retributive justice because the concept of blameworthiness is out of touch with scientific reality. This Note posits a more practical reason for reform by reviewing available empirics on the way people perceive human agency. The research suggests that as the science of human agency becomes increasingly vivid and reductionistic, laypeople will become proportionally less willing to attribute blame, and these shifting societal intuitions will ultimately diminish criminal law's moral credibility. The practical effects of low moral credibility might include diminished compliance, cooperation, and acquiescence with criminal laws, as well as increased general deviance. Importantly, this Note observes that these effects will likely manifest even if people retain a belief in free will. Further, ontological reality plays no part in this Note's argument; whether we in fact have free will is irrelevant. This Note instead contributes to the discourse by highlighting the implications of oncoming shifts in lay conceptions of both particular behaviors and the natural world writ large.

---

* J.D., 2015, University of Virginia School of Law; B.A., 2012, College of William & Mary. I am indebted to Professors Barbara Spellman and Richard Bonnie for their helpful insights, guidance, and criticisms; Professor Paul Sheldon Davies, whose seminars inspired this Note; the editorial staff of the *Yale Journal of Health Policy, Law & Ethics* for their input and thoughtful review; and Ben Carper, for being the sounding board against whom ideas herein reverberated.

TABLE OF CONTENTS

TABLE OF CONTENTS **290**

INTRODUCTION **291**

I. REFORM VS. CONSERVATION: THE CURRENT STATE OF THE DISCOURSE **298**

A. FREE WILL'S PLACE IN CRIMINAL LAW 299

B. THE DEBATE: REFORMIST ARGUMENTS AND CONSERVATIONIST RESPONSES 303

C. THIS NOTE'S ROLE IN THE DISCOURSE 306

II. MORAL CREDIBILITY: THE PRACTICAL EFFECTS OF PERCEIVED INJUSTICE **308**

III. A LOOMING MORAL CREDIBILITY PROBLEM **311**

A. FOLK INTUITIONS: FACTS AND DETERMINANTS 312

B. THE NEW SCIENCE OF HUMAN AGENCY 318

C. BRINGING IT ALL TOGETHER: TWO DIMMERS AND A FLIP SWITCH 321

CONCLUSION **323**

INTRODUCTION

"All theory is against the freedom of the will; all experience for it."[1] Samuel Johnson's centuries-old aphorism captures the core free will problem in a choice few words—that is, that although our decisions and actions truly feel willed, a scientific worldview presupposes that every event is wholly the product of mechanistic, observable causes that seem to preclude a transcendent moment of agency.[2] Despite its elegance, Johnson's aphorism invites the same pejorative question as the free will debate writ large: who cares? Although laypeople are largely aware of the fate-versus-free-will distinction and have their own views on human agency,[3] the meat of this debate, like so many other topics in ontology and metaphysics, can be criticized as "unreal, impractical, or unimportant."[4] Aside from the occasional news article[5] or television program,[6] daily life rarely prompts people to ponder whether their actions are willed or determined by a complex matrix of past causes (an ontological theory called "determinism"),[7] and the arguments of the philosophers and theorists who do take the question seriously seem to echo in the closed loop of academia, not a legislative hall. Thus, the word "theory" in Johnson's aphorism may carry an unintended descriptive load: the free will debate is too conceptual to change the way we conceptualize, engage with, and structure our society.

Recent discoveries in neuroscience, genetics, biology, and the behavioral sciences, however, have shaken the dust off the free will debate and thrust it into the popular fore by painting vivid pictures of the causal mechanisms that drive us

---

1 PAUL SHELDON DAVIES, SUBJECTS OF THE WORLD; DARWIN'S RHETORIC AND THE STUDY OF AGENCY IN NATURE 137 (2009) (quoting JAMES BOSWELL, THE LIFE OF SAMUEL JOHNSON 291 (1924)).

2 *See id.* at 139.

3 *See* 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, 474-75 (2013) (reporting the results of studies showing that people tend to think human decisions are "undetermined by prior causes" and that our "universe [is] indeterministic rather than deterministic").

4 John L. Hill, Note, *Freedom, Determinism, and the Externalization of Responsibility in the Law: A Philosophical Analysis*, 76 GEO. L.J. 2045, 2045 (1988).

5 *See, e.g.*, David Eagleman, *The Brain on Trial*, ATLANTIC (June 7, 2011, 4:58 PM), http://www.theatlantic.com/magazine/archive/2011/07/the-brain-on-trial/308520; Tania Lombrozo, *Blame Your Brain: The Fault Lies Somewhere Within*, NPR (June 16, 2014, 2:30 PM), http://www.npr.org/blogs/13.7/2014/06/16/322556750/blame-your-brain-the-fault-lies-somewhere-within; Dennis Overbye, *Free Will: Now You Have It, Now You Don't*, N.Y. TIMES (Jan. 2, 2007), http://www.nytimes.com/2007/01/02/science/02free.html.

6 Jim Fallon, *Exploring the Mind of a Killer*, TED (July 16, 2009), https://www.ted.com/talks/jim_fallon_exploring_the_mind_of_a_killer/transcript?language=en.

7 *See* Hill, *supra* note 4, at 2049 (defining determinism).

to act.[8] Whereas past studies on uncontrolled determinants of human behavior were largely correlational,[9] new research on the neurological, genealogical, and endocrinal underpinnings of conscious decisionmaking have illuminated an increasing number of links in the causal chain of given behaviors.[10] These discoveries, which this Note collectively terms "the new science of human agency,"[11] are making it increasingly apparent that certain behaviors are the product of biological processes over which we have no control. In short, the second half of Johnson's aphorism—that "experience [is] for" free will—is losing ground.[12]

Consider the following three findings that, by one neuroscientist's account, evidence and exemplify a recent trend in neuroscience that will force us to "challenge our sense of self."[13] First, there is some evidence to suggest that those who feel they were born the wrong sex exhibit the neurophysiology of the sex with which they identify. In particular, in these individuals' brains, a certain nucleus within the hypothalamus whose size is sex-specific is the right size according to their gender identity, but the wrong size according to their chromosomes, organs, hormones, and other phenotypical traits.[14] Second,

---

8 *See* Stephen Morse, *Neuroscience and the Future of Personhood and Responsibility*, *in* CONSTITUTION 3.0: FREEDOM AND TECHNOLOGICAL CHANGE 113, 115, 121 (Jeffrey Rosen & Benjamin Wittes eds., 2011) (asserting that the deterministic premise that humans are "victims of neuronal circumstances" has begun to exert "a strong pull on the popular, educated imagination").

9 *See, e.g.*, Deborah W. Denno, *Human Biology and Criminal Responsibility: Free Will or Free Ride?*, 137 U. PA. L. REV. 615, 619-49 (1998) (summarizing research on the predictive value of genetics, hormones, neurophysiology, intellect, and sociological influences).

10 *See* Atiq, *supra* note 3, at 456-58 (describing a confluence of neuroscientific, genetic, and sociological research that, when combined, provide both "causal explanation and data on mechanism[s]" underlying antisocial behavior); Robert M. Sapolsky, *The Frontal Cortex and the Criminal Justice System*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y LONDON 1787, 1787 (2004) ("[N]eurobiology is beginning to provide the first hints of mechanistic explanations for our personalities, propensities and passions."); Eagleman, *supra* note 5.

11 For a full explanation of this term, *see infra* Section III.B.

12 *See* DAVIES, *supra* note 1, at 137-69 (arguing, with far more elegance and persuasive force than this parenthetical can hope to capture, that Johnson's aphorism is collapsing under the weight of scientific discoveries which collectively indicate that "we are blind to the nonconscious capacities of our minds that generate in us the illusions regarding our agency").

13 *See* Sapolsky, *supra* note 10, at 1787.

14 *Id.* This description summarizes a single study conducted in 2000 with findings largely limited to "male-to-female" transgender subjects. Frank M. Krujiver et al., *Male-to-Female Transsexuals Have Female Neuron Numbers in a Limbic Nucleus*, 85 J. CLINICAL ENDOCRINOLOGY & METABOLISM 2034, 2034 (2000). The truncated description of transgender individuals above is meant to parallel researchers' description of that subgroup as those who "experience themselves as being of the opposite sex, despite having the biological characteristics of one sex." *Id.* More recent scientific research into biological explanations

scientists can eliminate sexual promiscuity in certain mammalian species—in other words, change them from polygamous to monogamous—through gene therapy that enhances certain neurochemical reward receptors in the nucleus accumbens.[15] Third, Huntington's disease, a neurodegenerative disorder whose early symptoms can include increased aggression, hypersexuality, poor social judgment, and impulsivity, results from the mutation of one specific gene among tens of thousands.[16] A common thread uniting these three examples is that each demystifies, to some degree, a particular personality trait, propensity, or behavior by elucidating one or more of its neurological and biological substrates. If the human affect is a car, the new science of human agency is a mechanic who has popped the hood to see how she runs.

Although these and similar findings fail to provide complete explanations for why different behaviors and propensities arise, the fraction of the causal matrix they do clarify has inspired some to reconsider popular notions of moral responsibility.[17] That is, if human thoughts and actions are the product of uncontrolled causes that include biological and neural mechanisms, genetic predispositions, and past experiences (the basic premise of determinism),[18] do we still deserve credit or blame for our actions? Attempted answers to this question vary widely, but generally fall into two categories. Those who answer "yes" call themselves "compatibilists" to reflect the idea that causally determined actions

---

behind gender identity continues to suggest that a detectable correlation may exist between neurological morphology and experienced gender. *See* Francine Russo, *Is There Something Unique About the Transgender Brain?*, SCI. AM. MIND (Jan. 1, 2016), http://www.scientificamerican.com/article/is-there-something-unique-about-the-transgender-brain (summarizing recent studies finding differences in the subcortical, cortical, and steroid-response features of the brain).

15 *See* Sapolsky, *supra* note 10, at 1788. Again, more recent studies have further suggested that differences in neurochemical signaling within the nucleus accumbens can cause intra- and interspecies variation in social behavior. *See* Alaine C. Keebaugh et al., *RNAi Knockdown of Oxytocin Receptor in the Nucleus Accumbens Inhibits Social Attachment and Parental Care in Monogamous Female Prairie Voles*, 10 SOC. NEUROSCIENCE 561, 566 (2015).

16 *See* Sapolsky, *supra* note 10, at 1787.

17 *See, e.g.*, Paul Bloom, *Free Will Does Not Exist. So What?*, CHRON. REV. (Mar. 18, 2012), http://chronicle.com/article/paul-bloom-free-will-does-not/131170; Hilary Bok, *Want To Understand Free Will? Don't Look to Neuroscience*, CHRON. REV. (Mar. 18, 2012), http://chronicle.com/article/hilary-bok-want-to-understand/131168; Jerry A. Coyne, *You Don't Have Free Will*, CHRON. REV. (Mar. 18, 2012), http://chronicle.com/article/jerry-a-coyne-you-dont-have/131165; Michael S. Gazzaniga, *Free Will Is an Illusion, but You're Still Responsible for Your Actions*, CHRON. REV. (Mar. 18, 2012), http://chronicle.com/article/michael-s-gazzaniga-free/131167.

18 *See* John Lawrence Hill, *Law and the Concept of the Core Self: Toward a Reconciliation of Naturalism and Humanism*, 80 MARQ. L. REV. 289, 330 (1997).

can be "free" under certain circumstances, such as when those actions result from conscious deliberation or rational thought.[19] Most theorists hold this "metaphysically modest" view of free will.[20]

So-called "incompatibilists," by contrast, point out that preserving a distinction between willed and unwilled behavior is purely semantic if our thoughts and behaviors—even conscious and rational ones—are the inevitable result of uncontrolled causes.[21] Thus, incompatibilists believe that free will and moral responsibility are impossible in a deterministic world.[22] A third worldview, "libertarianism," rejects determinism altogether and thus broadly preserves the possibility of free will.[23] Because libertarian theories posit that humans have a special capacity to transcend the natural world, academics largely reject libertarianism as a "metaphysically immodest conception of the human actor."[24] Among laypeople, however, libertarianism has considerable traction.[25]

The new science of human agency has sparked a closely related debate in legal academia—namely, whether a criminal justice system that justifies punishment in part on retributive grounds should change in light of empirical evidence casting doubt on the traditional notion that humans are self-causing agents.[26] This debate has focused on whether, as a theoretical matter,

---

19 *See* Gazzaniga, *supra* note 17.

20 Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, 359 PHIL. TRANSACTIONS ROYAL SOC'Y LONDON 1775, 1776-77 (2004) (describing "compatibilism" as the "dominant view among philosophers and legal theorists" and the official basis for "current legal doctrine").

21 *See* OWEN D. JONES ET AL., LAW AND NEUROSCIENCE 129 (1st ed. 2014). Put differently, incompatibilists decline to characterize rationally made, thoughtful decisions as "free" because, in a deterministic world, those underlying rationalizations and thoughts were themselves the product of past causes. Under this paradigm, then, there is no aspect of a given outcome that lacks a comprehensive set of causal forces whose net effect produced the behavior in question.

22 *Id.*

23 *Id.*

24 *See* Atiq, *supra* note 3, at 462-63; Galen Strawson, *The Impossibility of Moral Responsibility*, 75 PHIL. STUD. 5, 18 (1994).

25 *See* Atiq, *supra* note 3, at 479 ("[T]he folk concept of free will seems libertarian."); Anders Kaye, *Resurrecting the Causal Theory of the Excuses*, 83 NEB. L. REV. 1116, 1139 (2005); *cf.* Eddy Nahmias, *Folk Fears About Freedom and Responsibility: Determinism vs. Reductionism*, 6 J. COGNITION & CULTURE 215, 216 (2006) (paraphrasing experimental findings suggesting that "in certain conditions, most people express incompatibilist and libertarian intuitions"); Peggy Sasso, *Criminal Responsibility in the Age of "Mind-Reading,"* 46 AM. CRIM. L. REV. 1191, 1218 (2009) (same).

26 *See generally* Atiq, *supra* note 3, at 458-65 (describing the debate between reformists and their critics). Criminal punishment has four traditional justifications. In additional to retribution, these include deterrence, incapacitation, and rehabilitation. *See* Michele Cotton, *Back with a Vengeance: The Resilience of Retribution as an Articulated Purpose of Criminal*

foundational principles and specific doctrines of criminal law require the actors within its system to have free will.[27] For example, proponents of reform argue that because both the moral theories underlying the American legal system and specific rules like the voluntary act requirement, insanity defense, and other excuse defenses presuppose the existence of free will, the new science of human agency ought to drive revisions.[28] Critics of the reformists generally respond that free will is not foundational to criminal law and that, even if it is, the free will debate described above is far from settled.[29] This back-and-forth reveals that although the debate addresses important issues of fairness, egalitarian distribution of punishment, and civil liberty, it, much like the free will problem itself, is largely theoretical. Reform plainly has practical consequences, but both sides of the debate have failed to construct evidence-based arguments that their proposed course of action will in fact maximize the societal benefits of criminal law.[30]

For example, two particularly prominent reformists, Professors Joshua Greene and Jonathan Cohen, have made the controversial "empirical prediction"[31] that scientific advancements will gradually drive laypeople to doubt the "common sense, libertarian conception of free will and the retributivist thinking that depends on it."[32] This claim's persuasive value, however, is dubious in light of two objections. First, Green and Cohen fail to offer any evidence—for example, from surveys or the science of human cognition—that this widespread moral evolution will in fact occur (the "substantiation objection"). Second, Green

---

Punishment, 37 AM. CRIM. L. REV. 1313, 1313 (2000).

27 *See* Atiq, *supra* note 3, at 465.

28 *See id.* at 458-59.

29 *Id.* at 463-66.

30 *See infra* Section I.B (describing reformist arguments and conservationist responses).

31 *See* Greene & Cohen, *supra* note 20, at 1781.

32 *Id.* at 1776, 1781 (characterizing their argument as amounting to "an empirical prediction that . . . as more and more scientific facts come in, providing increasingly vivid illustrations of what the human mind is really like, more and more people will develop moral intuitions that are at odds with our current social practices"). Ideas akin to Greene and Cohen's have attracted considerable attention both inside and outside academia. In 2007, the John D. and Catherine T. MacArthur Foundation announced an initial investment of $10 million to fund research on the intersection between law and neuroscience. The fund's establishment was inspired in part by a building wave of both academic and mainstream literature, including Greene and Cohen's work, that, according to the Foundation, has upset centuries-old notions of human nature and posed an important question: "How would the law deal with theories that suggest that people's actions are not the direct result of prior intentions, that free-will is an illusion, that consciousness itself is a mere penumbra of the brain's activities?" Jonathan Fanton, President, John D. & Catherine T. MacArthur Found., Announcement of Law and Neuroscience Project (Oct. 9, 2007), https://www.macfound.org/press/speeches/announcement-law-and-neuroscience-project-jonathan-fanton-federal-court-house-new-york-ny-october-9-2007.

and Cohen fail to offer any evidence that the moral evolution will have any particular effects that suggest reform is in order (the "practicality objection"). For that reason, their argument amounts to an armchair prediction based on intuitive assumptions instead of empirical evidence.

This Note draws from three strands of research to address each shortcoming: studies on the moral credibility of criminal law, the folk psychology of free will and moral responsibility, and the new science of human agency. By considering these bodies of research in concert and comparing them to American criminal law doctrine, this Note assesses the empirical credence of Greene and Cohen's prediction and suggests its practical implications if true. Ultimately, this Note argues that lay perceptions of culpability will, in fact, adjust alongside the increasing degree to which natural, physiological explanations exist for given criminal behaviors. It next concludes that if the law fails to reflect these changing societal perceptions, the American justice system's regulatory strength will gradually erode.

The first strand of research, which relates to moral credibility, addresses the practicality objection by describing the concrete effects of perceived injustice. Studies on moral credibility suggest that when moral intuitions do not align with the moral principles reflected in criminal laws, society becomes incrementally less willing to acquiesce, assist, and defer not only to those same criminal laws, but also to completely unrelated laws.[33] Thus, available evidence suggests that if Greene and Cohen's prediction that scientific advancements will eventually undermine popular conceptions of free will is correct, the resultant clash between society's morals and retributive aspects of criminal law will diminish the latter's efficacy.

The second and third strands address the substantiation objection by showing that neuroscience and related fields are progressing in ways that have been shown to diminish people's penchant for retribution. The second strand, which examines lay intuitions about free will and moral responsibility, collectively reports that although people largely believe in a robust notion of free will,[34] they attribute less blameworthiness to criminals whose behaviors resulted from an obvious and specific set of causal antecedents.[35] The clearer the causal chain, the less culpable

---

33 *See, e.g.*, PAUL H. ROBINSON & JOHN M. DARLEY, JUSTICE, LIABILITY & BLAME: COMMUNITY VIEWS AND THE CRIMINAL LAW 6-7 (1995); Josh Bowers & Paul H. Robinson, *Perceptions of Fairness and Justice: The Shared Aims and Occasional Conflicts of Legitimacy and Moral Credibility*, 47 WAKE FOREST L. REV. 211, 258-62 (2012); Paul H. Robinson et al., *The Disutility of Injustice*, 85 N.Y.U. L. REV. 1940, 1995-97 (2010).

34 *See* sources cited *supra* note 25.

35 *See* ROBINSON & DARLEY, *supra* note 33, at 127-50; Sasso, *supra* note 25, at 1221; Azim F. Shariff et al., *Free Will and Punishment: A Mechanistic View of Human Nature Reduces Retribution*, 25 PSYCHOL. SCI. 1563, 1568 (2014); Azim F. Shariff & Kathleen D. Vohs, *The World Without Free Will: What Happens to a Society That Believes People Have*

an actor appears. The final strand suggests that recent studies, particularly in neuroscience, genetics, and sociology, are beginning to illuminate more and more segments of the causal matrix underlying given behaviors.[36] Thus, it appears that the new science of human agency will have downstream effects that ultimately support Greene and Cohen's calls for reform: an increasingly naturalistic picture of human behavior causes less retributivism; less retributivism diminishes moral credibility (assuming static criminal law doctrine);[37] and a widespread erosion of criminal law's moral credibility portends declining efficacy.

This Note proceeds by building this cascading chain of inferences in three Parts. Part I describes the current state of the debate on whether modern science should catalyze doctrinal changes and clarifies the discursive void this Note aims to fill. Part II responds to the practicality objection by describing how population-wide changes in conceptions of morality can incrementally weaken the effectiveness of the justice system. Part III argues that the new science of human agency presents a looming moral credibility problem for criminal law in light of likely shifts in the folk psychology of desert.

Before launching into discussion, however, I offer one last clarifying proviso critical to understanding this Note's logical structure: The argument that follows assesses the implications of human perception, not ontological reality. Put differently, all that matters for present purposes are people's views on free will and human responsibility, not whether free will and responsibility in fact exist. Accordingly, this Note neither defends determinism nor depends on it as a logical premise. Instead, this Note constructs an evidence-based objection to retributivism using available empirics that collectively reveal the striking malleability of human blame attributions, some of the particular situational forces

---

No Conscious Control Over Their Actions?, 6 SCI. AM. 77, 78 (2014); Vincent Yzerbyt & Anouk Rogier, *Blame It on the Group: Entitativity, Subjective Essentialism, and Social Attribution*, *in* THE PSYCHOLOGY OF LEGITIMACY: EMERGING PERSPECTIVES ON IDEOLOGY, JUSTICE, AND INTERGROUP RELATIONS 103, 123-24 (John T. Jost & Brenda Major eds., 2001).

36 *See, e.g.*, Atiq, *supra* note 3, at 454-57; Sapolsky, supra note 10, at 1794; *see generally* DAVID M. EAGLEMAN, INCOGNITO: THE SECRET LIVES OF THE BRAIN (2011) (describing a number of unconscious and nonconscious substrates of human behavior and decisionmaking); JAMES FALLON, THE PSYCHOPATH INSIDE: A NEUROSCIENTIST'S PERSONAL JOURNEY INTO THE DARK SIDE OF THE BRAIN 9 (2013) ("In my mind, we are machines, albeit machines we don't understand all that well, and I have believed for decades that we have very little control over what we do and who we are. To me, nature (genetics) determines about 80 percent of our personality and behavior, and nurture (how and in what environment we are raised) only 20 percent.").

37 Commentators often bemoan the seemingly glacial pace of criminal law's response to innovation—both scientific and otherwise. *See, e.g.*, Richard Addelstein, *Victims as Cost-Bearers*, 3 BUFF CRIM. L. REV. 131, 169 (1999); Michael Rustad, *Private Enforcement of Cybercrime on the Electronic Frontier*, 11 S.C. INTERDISC. L.J. 63, 96 (2001); *infra* note 41.

that affect our willingness to punish others, and the practical effects that rear when the law punishes people in ways we do not think it should. While some theorists have used the new science of human agency to argue that free will is an illusion and retributivism is unjust,[38] this Note repurposes that same science in service of a comparatively modest project: assessing both the likelihood[39] and implications[40] of a wide-scale recalibration of society's moral compass. Using that approach, this Note intends to accomplish something the first paragraph of this Introduction dismissed as unlikely: show that the free will debate has practical import.

I. REFORM VS. CONSERVATION: THE CURRENT STATE OF THE DISCOURSE

Although commentators and theorists often decry the legal system's incessant failure to adapt to relevant scientific developments,[41] evidence-based changes do occur on occasion. Consider the example of eyewitness experts.[42] In light of scientific evidence suggesting that eyewitness identification evidence is categorically unreliable and responsible for a significant number of wrongful convictions, many jurisdictions have begun allowing expert witnesses to help the jury properly weigh identification testimony by summarizing the helpful science in court.[43] In this example, the relevant science and attendant legal changes were both somewhat narrow in scope.

This Part describes a far more ambitious call for reform based on an ontological proposition synthesized from a vast array of scientific findings. A summary of these reformists' charge is as follows: Because the new science of human agency calls into question traditional conceptions of human responsibility, aspects of criminal law based on those outdated conceptions are unjustifiable.[44]

Discussion in this Part proceeds by first describing aspects of the criminal law that reformists aim to change—namely, criminal law's foundational

---

38 *See infra* Section I.B.

39 *See infra* Part III.

40 *See infra* Part II.

41 *See, e.g.*, Viktoras Justickis, *Does the Law Use Even a Small Proportion of What Legal Psychology Has To Offer?*, *in* PSYCHOLOGY AND LAW: BRIDGING THE GAP 224, 225-27 (David Canter & Rita Zukauskiene, eds., 2008) (asserting that "psychology rarely informs the law in practice"); Sapolsky, supra note 10, at 1788 ("[T]here are an ever-increasing number of realms in which the legal system has made little headway incorporating neurobiology.").

42 *See generally* Brian L. Cutler & Gary L. Wells, *Expert Testimony Regarding Eyewitness Identification*, *in* PSYCHOLOGICAL SCIENCE IN THE COURTROOM: CONSENSUS AND CONTROVERSY 100 (Jennifer L. Skeem et al. eds., 2009) (arguing for the use of expert witnesses to evaluate the appropriate credibility of eyewitnesses' testimony).

43 *Id.*

44 *See, e.g.*, Greene & Cohen, *supra* note 20, at 1776; Sapolsky, *supra* note 10, at 1788; Eagleman, *supra* note 5.

dependence on retributivism, as well as specific doctrines and defenses that in some way contemplate the offender's capacity for volition. Next, it describes the positions of reformists, who feel that current law ought to change in light of the new science of human agency, as well as responses from conservationists, who argue the opposite. This Part concludes by identifying a gap in the literature that this Note aims to fill.

A. Free Will's Place in Criminal Law

Because reformists contend that the new science of human agency calls into question common attributions of blame and moral responsibility, their proposals for reform present radical changes implicating both criminal law's foundational premises and specific criteria for criminal responsibility.[45] Although part of the debate involves different interpretations of the criminal law as written, some of the criminal law's underlying premises and specific doctrines are uncontroversial. Both sides agree that black-letter tenets of the criminal code largely do not allow defendants to use the new science of human agency to construct "volitional impairment" defenses—for example, that the criminal is not blameworthy because his criminal conduct was the product of a genetic predisposition triggered by environmental stimuli. Criminal law, as currently structured, largely precludes these sorts of defenses because it assumes human actors have some form of free will, a foundational premise that permeates specific doctrines.

Blame, responsibility, and free will play preeminent roles in the criminal justice system; these concepts are in some sense "foundational."[46] Indeed, this idea is hornbook law:

> The criminal law is based on the capacity of the individual to make free choices and the assumption that virtually all of our behavior virtually all of the time is a result of free choice. This may or may not be a description of reality. But the criminal law as we know it cannot function without the hallmarks of responsibility, blame, and punishment as the working premises for most behavior.[47]

Thus, instead of construing human agents in ways that are rigorously naturalistic and scientific, criminal law operates by presuming we all have some

---

45 *See* Atiq, *supra* note 3, at 458-60.

46 *Id.* at 458, 465; *see* Greene & Cohen, *supra* note 20, at 1783 (reformist); Morse, *supra* note 8, at 123-24 (conservationist); Sapolsky, *supra* note 10, at 1793 (reformist); Eagleman, *supra* note 5 (reformist).

47 PETER W. LOW, BLACK LETTER OUTLINES: CRIMINAL LAW 199 (3d ed. 2007).

capacity for free will and moral responsibility that might justify retributive punishment.[48] Accordingly, Professor Stephen Morse characterizes criminal law's view of the human actor as "folk psychological" in that, in the law's eyes, criminal behavior must be describable using the presence or lack of colloquial, familiar mental states (for example, murderous rage or a premeditated intent to kill).[49] Although this folk psychological view of the human actor is arguably "primitive [and] pre-scientific,"[50] it has endured as a descriptive tenet of the criminal law's current organization. No matter what scientists and philosophers say about our volitional capacities, the law presumes that humans have free will in some sense of the term.[51]

This underlying assumption—that humans have "the general capacity for rationality" and the ability to "understand the good reasons for action and . . . conform[] to legal requirements through intentional action or forbearance"[52]—permeates specific criminal law doctrines, including the voluntary act requirement, required culpable mental states, and defenses of excuse.[53] As a result, it is difficult under current doctrine to earn acquittal using a volitional impairment defense under any of those three theories.[54]

Take, for example, the insanity defense, which "has traditionally been understood as vindicating the free will assumption"[55] and rarely succeeds. The insanity defense has two variations—the cognitive dysfunction test and the control dysfunction test. Typical formulations of the cognitive dysfunction test require the defendant to show that he could not appreciate the wrongfulness of his action.[56] The control dysfunction (or "irresistible impulse") test, by contrast,

---

48 *See* Morse, *supra* note 8, at 127 ("At present, the law's official position [is that] conscious, intentional, rational and uncompelled agents may properly be held responsible . . . ."); *see also* State Farm Fire & Cas. Co. v. Brown, 905 P.2d 527, 535 (Ariz. Ct. App. 1995) ("The legal model's postulate of free will envisions people as morally and legally answerable for their conduct rather than as pigeons in a Skinner box. By contrast, the scientific model in most schools of psychology is largely deterministic . . . .").

49 *See* Morse, *supra* note 8, at 127.

50 *Id.* at 124.

51 *Id.* Note the significance of the phrase "in some sense of the term." Reformists argue that criminal liability depends on a libertarian notion of free will, which involves a metaphysically robust moment of agency in which the agent transcends the laws of nature as a true first causer. Conservationists, by contrast, argue that criminal law only requires a compatibilist ontology in which "free" acts are those that are the product of conscious, rational deliberation.

52 *Id.* at 125.

53 *See* Atiq, *supra* note 3, at 459-60; Morse, *supra* note 8, at 124-25.

54 *See* Atiq, *supra* note 3, at 457-58.

55 Michele Cotton, *A Foolish Consistency: Keeping Determinism out of the Criminal Law*, 15 B.U. PUB. INT. L.J. 1, 5 (2005).

56 *See* ROBINSON & DARLEY, *supra* note 33, at 129; *see, e.g.*, MODEL PENAL CODE §

enables a defendant to plead insanity even if he understood the nature of his actions so long as he lacked the ability to control his conduct and his free agency was "destroyed."[57] Jurisdictions differ on whether defendants may plead an insanity defense under one or both of these theories of insanity.[58] Since the illustrious John Hinckley trial and the resultant passage of the Federal Insanity Defense Reform Act, which aimed in part to shrink the insanity defense's scope and availability, the number of jurisdictions that allow defendants to plead control dysfunction has shrunk to nineteen.[59] Thirty jurisdictions allow only cognitive dysfunction and two allow no insanity defense at all.[60] Further, regardless of the type of dysfunction pleaded, successful insanity defenses are quite rare; one recent study indicated that defendants plead insanity in under 1% of criminal indictments and that insanity defenses fail 71% of the time.[61] Thus, American criminal codes largely reject the idea that a person whose actions were entirely the product of uncontrolled mental and sociological processes should earn acquittal by virtue of insanity, and in the minority that do, the chances of succeeding on such an argument are slim.

"Diminished capacity" defenses have proved similarly unsuccessful. Under this strategy, defendants use volitional impairment evidence to negate mens rea

---

4.01(1) (AM. LAW INST., Official Draft 1985) ("A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity . . . to appreciate the criminality [wrongfulness] of his conduct . . . ."); Clark v. State, 588 P.2d 1027, 1029 (Nev. 1979) (noting that the jury's duty under the M'Naughten rule was to determine whether the "appellant knew the nature and quality of her acts, had the capacity to determine right from wrong or knew whether she was doing wrong when she committed the crime").

57 *See* ROBINSON & DARLEY, *supra* note 33, at 129-30; *see, e.g.*, MODEL PENAL CODE § 4.01(1) ("A person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity . . . to conform his conduct to the requirements of the law."); Godley v. Commonwealth, 343 S.E.2d 368, 370 (Va. Ct. App. 1986) ("The defense is only available where the accused's mind has become 'so impaired by disease that he is totally deprived of the mental power to control or restrain his act.'") (quoting Thompson v. Commonwealth, 70 S.E.2d 284, 292 (Va. 1952)).

58 *See* Robinson et al., *supra* note 33, at 1956.

59 Melinda Carrido, Note, *Revisiting the Insanity Defense: A Case for Resurrecting the Volitional Prong of the Insanity Defense in Light of Neuroscientific Advances*, 41 SW. U. L. REV. 309, 311, 319-22 (2012) (discussing the Insanity Reform Defense Act, Pub. L. No. 98-473, § 402(a), 98 Stat. 2057, 2057 (1984) (codified as amended at 18 U.S.C. § 17(a) (2012))). Passage of the Insanity Defense Reform Act reflected, to some degree, Congress's response to the national furor that accompanied Hinckley's acquittal. See United States v. West, 962 F.2d 1243, 1249 (7th Cir. 1992); United States v. Cameron, 907 F.2d 1051, 1061 (11th Cir. 1990).

60 Robinson et al., *supra* note 33, at 1955-56. These numbers are accurate as of 2010.

61 Cotton, *supra* note 55, at 18 & n.90.

elements.[62] As a practical matter, Federal Rule of Evidence 704(b) and state law equivalents limit this approach by preventing experts from testifying that a particular mental disease or defect negated a defendant's culpable mental state.[63] Thus, rules like 704(b) preclude experts from entering into evidence the deterministic hypothesis that, although a given defendant's actions meet the relevant actus reus elements, he lacked the requisite mental state—purpose, knowledge, recklessness, or negligence[64]—because his conduct was instead the product of subconscious, nonconscious, or otherwise uncontrolled causes. Note that this hypothesis is contrary to the folk psychological conception of human acts, which posits that all behavior is "at least rationalizable by mental-state explanations."[65]

The few innovative defense lawyers that have sidestepped 704(b) and attempted this strategy have found mixed success.[66] For example, in a 2007 sexual abuse case, the defendant admitted evidence that frontal lobe defects rendered him physiologically unable to form the sort of intent or plan required under the relevant statute.[67] That defendant was convicted and received a sentence of eighteen years to life.[68] In the notorious "twinkie defense" case, by contrast, a defendant successfully avoided a first degree murder conviction by arguing that a combination of junk food and extreme stress altered his mental state at the time of the killings.[69] Although these sorts of biological deficiency cases are highly publicized, they are quite risky and exceedingly rare.[70]

As currently structured, the American criminal justice system offers defendants limited means to avoid or diminish criminal liability through evidence that their criminal behavior was causally determined by biological,

---

62 *See id.* at 18-23; *see also* Deborah W. Denno, *Crime and Consciousness: Science and Involuntary Acts*, 87 MINN. L. REV. 269, 285 (2002) ("Diminished capacity can either be a complete defense resulting in an acquittal, like automatism/unconsciousness, or, more commonly, a partial defense resulting in the defendant's conviction of a lesser crime.").

63 *See* FED. R. EVID. 704(b); Cotton, *supra* note 55, at 19-20 & n.95 (listing similar state rules).

64 These four culpable mental states were drawn from the Model Penal Code provision that details general culpability requirements. *See* MODEL PENAL CODE § 2.02(2) (AM. LAW INST., Official Draft 1985).

65 Morse, *supra* note 8, at 123.

66 *See* Denno, *supra* note 9, at 616 (noting a series of innovative defenses in recent, high-profile criminal cases).

67 *See* Atiq, *supra* note 3, at 457.

68 *Id.* at 458.

69 Denno, *supra* note 9, at 616-17 (referring to Dan White's diminished-responsibility defense to first-degree murder charges for the killing of Mayor George Moscone and Supervisor Harvey Milk).

70 *Id.* at 616 ("Judges and juries have not accepted most of the 'new' and highly publicized criminal law defenses.").

neurophysiological, psychological, and environmental causes. To wit, neither the insanity defense nor the diminished capacity defenses offer defendants reliable means to achieve those ends. This should come as no surprise given that, as the Supreme Court has noted, "a deterministic view of human conduct . . . is inconsistent with the underlying precepts of our criminal justice system."[71]

One set of these "underlying precepts" that determinism does not offend, however, is our criminal justice system's triad of utilitarian goals: deterring future crime (deterrence), eradicating recidivism (rehabilitation), and removing irreversibly dangerous criminals from society (incapacitation).[72] Unlike our system's fourth traditional goal—retribution—utilitarian goals aim only to achieve practical benefits for society without labeling criminal misconduct as morally right or wrong.[73] For that reason, a hypothetical system founded only on utilitarian ideals would not crumble upon the wholesale rejection of free will and responsibility as accurate models of human behavior. In that system, each blameless violator would receive punishments on the sole rationale that those punishments affect behavior in ways that benefit society. Thus, although certain aspects of the American criminal justice system currently depend on the concepts of free will and responsibility, not all do. The question thus becomes: Given that modern science is gradually lending stronger and stronger empirical credence to a deterministic model of the human actor, should those aspects of our system dependent upon concepts of free will and responsibility adapt?

B. The Debate: Reformist Arguments and Conservationist Responses

The charge for reform reduces to two principal arguments, one normative and the other predictive: first, that current legal doctrine is unjustifiable insofar as it assumes a libertarian notion of free will that contemporary science overwhelmingly rejects;[74] and second, that the criminal justice system reflects societal intuitions of justice, which will soon evolve towards a deterministic worldview in light of compelling scientific discoveries.[75] These two arguments are related; in theory, as contemporary science convinces more and more people of the normative proposition, the criminal law will adapt as suggested in the predictive one.

---

71 United States v. Grayson, 438 U.S. 41, 52 (1978). This conjecture is either poorly worded or evidence that the *Grayson* majority are incompatibilists. As noted above, compatibilists believe that free will and blameworthiness are intelligible concepts in a deterministic world.

72 Cotton, *supra* note 26, at 1316-17.

73 *Id.*

74 *See, e.g.*, Sapolsky, *supra* note 10, at 1794; Eagleman, *supra* note 5.

75 *See generally* Greene & Cohen, *supra* note 20, at 1776 (asserting that the findings of modern science will cause widespread "rejection of free will").

Reformists' first argument reflects the general idea that because criminal law's bedrock principles of blame and responsibility are outdated, the law itself is unfair, inhumane, and needlessly inconsistent with scientific reality.[76] Neuroscientists David Eagleman and Robert Sapolsky each relay this view in articles similar in structure. Both describe especially vivid neurological mechanisms known to affect behavior before explaining why, in light of that science, the natural world is best described as deterministic and incompatible with moral responsibility.[77] Eagleman, for example, reviews the science of genetics, unconscious cognitive processes, and environmental influences before arguing that a "modern understanding of the brain" requires "[b]lameworthiness [to] be removed from the legal argot."[78] Instead of justifying punishment on the basis of retribution or desert, Eagleman argues, punishments should focus exclusively on the consequentialist, utilitarian goals of deterrence, rehabilitation, and incapacitation.[79]

Sapolsky's position is similarly incompatibilistic. He asserts that, given the growing body of research providing "mechanistic explanations for our personalities, propensities, and passions," including evidence that a malfunctioning prefrontal cortex renders some individuals biologically incapable of making "good" decisions instead of "bad" ones, the law's focus on blame instead of past causes is misguided.[80] Accordingly, just like Eagleman, Sapolsky recommends dispensing with retributivism, albeit in a more colorful fashion:

> To understand is not to forgive or to do nothing; whereas you do not ponder whether to forgive a car that, because of problems with its brakes, has injured someone, you nevertheless protect society from it. . . . [And] although it may seem dehumanizing to medicalize people into being broken cars, it can still be vastly more humane than moralizing them into being sinners.[81]

Whereas the neuroscientists' argument is normative, the legal theorists' argument is perhaps best understood as predictive. Professors Joshua Greene and Jonathan Cohen advance the provocative and much-discussed[82] view that although laypeople overwhelmingly believe themselves to have a robust power of free will, neuroscientific advancements will gradually change that intuition by revealing, with increasing clarity and vividness, that our actions are driven by

76 *See* sources cited *supra* notes 71-72.
77 *Id.*
78 Eagleman, *supra* note 5.
79 *Id.*
80 Sapolsky, *supra* note 10, at 1788.
81 *Id.* at 1794.
82 *See, e.g.*, Atiq, *supra* note 3, at 458.

neurological and environmental processes that we cannot control.[83] This moral revolution, they argue, will involve society's wholesale rejection of both free will and moral responsibility—that is, people will reject both libertarianism (robust free will) and compatibilism ("metaphysically modest"[84] free will)—and thus adopt a worldview that clashes with the law's current dependence on blame and retributivism.[85] Accordingly, because criminal law doctrines "exist because they more or less adequately capture an intuitive sense of justice," the law will inevitably evolve by replacing retributivist justifications for punishment with utilitarian ones.[86]

Critically, Greene and Cohen do not assert that criminal law "officially" depends on robust free will.[87] They admit that criminal law purports to only require a "metaphysically modest" version of free will that is compatible with both determinism and retribution.[88] But regardless of what the law "*says*,"[89] they argue, it *in fact* depends on the intuitions that society currently ascribes to it: libertarianism and compatibilism.[90] That is, although it is logically coherent for the law to use words like blame, responsibility, and just deserts even if the world is deterministic thanks to the philosophy of compatibilism, that is an "unstable marriage" because it is not intuitive and the law necessarily reflects societal intuitions.[91] Simply put, society does not see things that way, so the law does not work that way.[92] Instead, society intuitively believes criminal law punishes people because they deserve it, an intuition that will lead people to question criminal law once the science convinces them that determinism is true and blame is nonsensical.

Conservationists like Professor Stephen Morse respond with two principal objections: (1) neuroscience does not currently support the conclusion that we are not agents; and (2) even if it does, criminal law need not change because it embraces a form of free will that is compatible with determinism.[93] First, Morse argues that because the science of agency is still "in its infancy,"[94] we cannot

---

83 *See* Greene & Cohen, *supra* note 20, at 1776.

84 *Id.* (advocating instead for a consequentialist view of free will).

85 *See supra* note 20 and accompanying text.

86 Greene & Cohen, *supra* note 20, at 1776.

87 *Id.*

88 *Id.* (defining this view as "compatibalism").

89 *Id.*

90 *Id.* ("[W]e argue that the law's intuitive support is ultimately grounded in a metaphysically, overambitious, libertarian notion of free will that is threatened by determinism and, more pointedly, by forthcoming cognitive neuroscience.").

91 *Id.*

92 *Id.*

93 *See* Morse, *supra* note 8, at 119-21.

94 *Id.* at 119.

justifiably claim that the mental states posited by the folk psychological model are "chimera[s]" that have zero explanatory power.[95] At present, neuroscience cannot yet explain "how molecules, which have no intentionality or temporal sense, produce intentional creatures with a sense of past, present and future that guide our lives."[96] Watching neurons fire on a brain scanner is one thing; providing a complete, physicalist explanation for a mental state (for example, someone's plan to steal a jewel) is wholly another. Because current science explains so little about these "brain-mind" and "brain-action" connections, Morse argues, reshaping our societal institutions based on it would amount to "neuroarrogance."[97]

In addition to this critique of the science, Morse emphasizes that which Greene and Cohen concede: substantive legal doctrine does not depend on robust free will.[98] That is, criminal law does not require human agents to all be transcendent self-causers in order to justifiably attribute blame. Criminal law instead operates on compatibilist premises that allow the system to attribute moral responsibility to any criminal actor with the capacity for "conscious, intentional, [and] rational" behavior.[99] We can still distinguish between the folk psychological states of conscious behavior and unconscious or uncontrolled behavior even in a deterministic world. Thus, at least in theory, criminal law need not adapt if determinism is true.

C. This Note's Role in the Discourse

Morse's response to Greene and Cohen's empirical prediction is cogent but perhaps incomplete. He argues that because science might never disprove human agency, Greene and Cohen's envisioned moral revolution will not occur.[100] Conservationists might object to Greene and Cohen's prediction on two additional grounds.

First, Greene and Cohen fail to provide any evidence from the behavioral sciences that society will completely shed a popular, prevailing belief in robust free will, moral responsibility, and retributivism. In fact, the only scientific evidence they do offer is research from neuroscience and cognitive genetics

---

95 *Id.* at 122.

96 *Id.* (criticizing the post-Enlightenment "reductionist" view of free will).

97 *Id.* Since this Note focuses on the significance of human perception instead of the ontological validity of determinism, this Note takes no position on this aspect of the debate.

98 *Id.* at 119 ("[F]ree will plays no doctrinal role in criminal law and it is not genuinely foundational for criminal responsibility. Nor is determinism inconsistent with the folk psychological view of the person.").

99 *Id.* at 120.

100 *Id.* at 128.

indicating that humans are hardwired to punish others.[101] Thus, it is doubtful from their account that widespread beliefs in moral responsibility will erode even in the face of scientific advancements that support a deterministic worldview. Recall that this Note termed this apparent unlikelihood the "substantiation objection."

Second, assuming this moral revolution does occur, its practical implications are unclear. Greene and Cohen assert that the law will simply adapt to align with prevailing societal views, but this account is opaque and largely unsupported in their article. Not only do the authors fail to specify how the law will adapt, but also why—legislators often pass laws, policies, and programs that people disagree with on some level. Greene and Cohen fail to offer practical reasons why we should completely reorganize our criminal law to accommodate people's feelings aside from, perhaps, the ethereal promise of fairness. Recall that this Note termed this the "practicality objection."

This Note offers responses to both objections. First, it responds to the practicality objection by pointing to research on the moral credibility of substantive criminal laws and their real-world outputs.[102] Because studies on this topic indicate that perceived injustice incrementally diminishes people's willingness to comply with the rule of law, widespread rejection of one of criminal law's foundational tenets could weaken the law's regulatory efficacy.

Second, this Note predicts that advancements in neuroscience and related fields may in fact present a looming moral credibility problem in light of cutting-edge research showing that people—even those who believe in robust free will—tend to view defendants as less culpable if a granular and reductionistic explanation exists for their conduct.[103] Because the new science of human agency is beginning to illuminate these compelling physical explanations, people's moral intuitions about given defendants' culpability may in fact change *even if their views on free will remain static*.

That is a critical distinction worth briefly highlighting. Greene and Cohen argue that the new science of human agency will cause broad-based changes in criminal law once people stop believing in free will. This Note, by contrast, argues that the new science will have the practical effect of diminishing criminal law's efficacy in the near term even if people hold fast to libertarianism.

Finally, a quick word on this Note's discursive potency: this Note does not purport to comprehensively argue that, as a normative matter, the criminal law should dispense with retributivism or otherwise reorganize to accommodate changing communal views. Instead, it is best construed as a single arrow in the

---

101 Greene & Cohen, *supra* note 20, at 1784 (suggesting that "the impulse to exact punishment may be driven by phylogentically old mechanisms in the brain").

102 *See infra* Part II.

103 *See infra* Part III.

reformists' quiver; an evidence-based prediction that the new science of human agency may have important, practical implications for the criminal justice system even if Greene and Cohen's prediction of a general moral revolution in criminal law proves false.

II. MORAL CREDIBILITY: THE PRACTICAL EFFECTS OF PERCEIVED INJUSTICE

Greene and Cohen's empirical prediction that community views will soon clash with criminal law's bedrock principles of blame and responsibility has a number of commonsense implications.[104] Greene and Cohen foresee one such effect—that changes in social morality will precipitate changes in the law—but there are surely others. For example, in theory, criminal laws that accurately reflect society's moral intuitions about justice should constitute potent deterrents for anyone who values social acceptance.[105] A criminal code that fails to wield the power of our natural aversion to judgment, stigmatization, and interpersonal ostracism misses out on a cost-free, powerful guarantor of compliance.[106] Similarly, and perhaps more obviously, people are probably more likely to comply with laws they agree with simply because they think complying is the right thing to do.[107] These intuitions provide the beginnings of an answer to the practicality objection, but recent empirics indicate we can do better.

A growing body of research suggests that aligning the law's various moral judgments with those of society has a number of utilitarian benefits.[108] Researchers generally refer to the capacity of a given law to accurately and authoritatively reflect the moral intuitions of the relevant community as the law's "moral credibility."[109] Studies have shown that consequences exist for drafting or maintaining laws that lack moral credibility. When people disagree with a principle of justice that a given law reflects, they are less likely to comply with that law,[110] comply with other unrelated laws,[111] and cooperate with the criminal justice system as discretionary actors (for example, as witnesses and jurors).[112] Moral mismatches have also been shown to encourage deviant behavior in

104 See ROBINSON & DARLEY, *supra* note 33, at 6.

105 *See id.*

106 *See id.*

107 *See, e.g.*, *id.*

108 Robinson et al., *supra* note 33, at 1995.

109 ROBINSON & DARLEY, *supra* note 33, at 6; Bowers & Robinson, supra note 33, at 240-41; *see also* Robinson et al., *supra* note 33 (using "moral credibility" and "perceived justice" interchangeably).

110 *See* ROBINSON & DARLEY, *supra* note 33, at 7.

111 *See* Bowers & Robinson, *supra* note 33, at 262.

112 *See* ROBINSON & DARLEY, *supra* note 33, at 7.

general.[113] Further, the relationship between a law's moral credibility and its compliance power appears to be proportional.[114] Thus, people do not simply indicate they will comply with laws they agree with and flout the ones they do not. Instead, the extent of subjects' past transgressions and self-reported likelihood of future transgression tracks the depth of their moral objections to the laws they oppose.[115]

Like all correlational studies, however, studies on the effects of perceived injustice have limitations that affect their generalizability. Accordingly, to determine whether changes in societal intuitions about free will and moral responsibility will trigger the practical effects observed in the research, one must examine these studies' design and methodology in search of relevant constraints.

Studies on moral credibility have employed both controlled and observational research methods.[116] In the controlled studies, experimenters typically begin by exposing some subjects—only those in the experimental group—to a crime, law, or conviction that the subject may or may not characterize as just.[117] Methods of exposure include mock newspaper articles, mock television news reports, or simply telling the subject about the crime or legal outcome in an interview or questionnaire.[118] To gauge the effects of perceived injustice, experimenters either (1) observe subjects' frequency of engaging in subsequent deviant behavior like stealing a pen[119] or committing jury nullification in a mock trial;[120] or (2) rely on subjects' self-reported future likelihood of violating the law or cooperating with the criminal justice system in other ways.[121] Those other ways include reporting known crimes to authorities, turning in evidence to the police, and reporting their own accidental violations.[122]

Observational studies, by contrast, have the benefit of examining the effects of real-life exposure to injustice. Subjects in these studies served as jurors in criminal court proceedings,[123] committed a crime themselves,[124] or knew a friend

113 *See* Robinson et al., *supra* note 33, at 2011-16.

114 *Id*; *see also* Bowers & Robinson, *supra* note 33, at 258 (asserting that "[m]inor changes in moral credibility incrementally affect people's willingness to acquiesce, assist, and defer to the criminal law").

115 Robinson et al., *supra* note 33, at 2011.

116 *Id.*

117 *Id.*

118 *Id.*

119 *Id.* at 2015.

120 *Id.* at 2014.

121 *Id.* at 2011-16.

122 Bowers & Robinson, *supra* note 33, at 258.

123 *Id.* at 259.

124 Robinson et al., *supra* note 33, at 2012.

or relative who suffered a legal outcome that the subject perceived as unjust.[125] Thus, the behavioral aftereffects observed among these subjects might more accurately reflect reactions to moral objections.

Importantly, the results of observational and experimental studies align. Both indicate that maximizing the law's moral credibility brings the utilitarian benefits of improving criminal law's potency as a deterrent and increasing the likelihood that civilian actors will contribute to the efficacy of the system as jurors, witnesses, and bystanders.[126]

These findings have a number of limitations. First, the studies indicating that one law's low moral credibility subtly encourages people to flout other laws only observed or asked about future transgression of minor offenses—things like speeding, smoking marijuana, or parking illegally.[127] Thus, the extent of experimenters' observed "general deviance" effect is unknown.[128] Further, the studies' reliance on self-reporting to predict actual behavior may have been necessary for practical reasons, but is nevertheless dubious. Subjects may or may not be accurate judges of their own propensities, especially given that some may have been subconsciously primed to answer in certain ways due to the nature of preceding questions.

Finally, exposure—the method by which subjects learned that their version of justice conflicted with the law's—poses two separate limitations. First, the research does not indicate how likely it is that segments of the population who disagree with a given law will in fact confront that law. People must either experience moral mismatches directly (as a defendant or juror) or indirectly (through news reports or interpersonal contacts); low moral credibility does not per se diminish the law's efficacy. Because laypeople do not spend their Saturday mornings reading their state criminal codes, direct and indirect contact are probably the only two realistic mediums of exposure. Second, studies do not indicate how long a given person's exposure to a moral mismatch impacts their likelihood of future compliance. For all of these reasons, one must read the literature with caution.

Nevertheless, the science summarized above addresses the practicality objection by identifying certain effects that might follow shifts in society's moral landscape. To wit, the research indicates that (1) if a person observes or hears about a legal outcome she deems unjust, then (2) for an unknown period of time she will be less likely to comply, acquiesce, or cooperate with at least one

---

125 *Id.* at 2015-16.

126 ROBINSON & DARLEY, *supra* note 33, at 7.

127 *See, e.g.*, Robinson et al., *supra* note 33, at 2011-16.

128 *See* Marc Le Blanc & Rolf Loeber, *Developmental Criminology Updated*, 23 CRIM. & JUST. 115, 118-119 (1998) (discussing the definition and evolution of the term "general deviance" in criminology).

criminal law, procedure, or process. If Greene and Cohen's prediction is true—that society will soon adopt views on human agency that directly clash with criminal law's terms and consequences—then moral mismatches will result. Further, so long as individuals experience direct or indirect exposure to those terms and consequences, the research on moral credibility indicates that some percentage of those exposed will actually change their behavior in ways that conflict with our system's rules or impede its processes. The magnitude of this effect is currently unclear, however, because the research provides limited guidance on the extent to which populations will have direct or indirect exposure to legal outputs they disagree with, the length of time problematic behavior persists after exposure, and the number and nature of crimes that exposed individuals are likely to violate.

A conservative application of the research on moral credibility supplies a plausible reason for concern if the new science of human agency drives Greene and Cohen's predicted moral revolution and our criminal law remains unchanged. Further, because Greene and Cohen predict radical shifts in morality that could change society's holistic views of our criminal justice system, alternative applications of the science might predict more extreme effects. Thus, both conservative and aggressive extrapolations indicate that, at the very least, a gradual erosion of retributivism's moral credibility might result in practical effects worth consideration by policymakers: criminal law may begin to suffer diminishing compliance, cooperation, and efficacy.

III. A LOOMING MORAL CREDIBILITY PROBLEM

Part II assumes that Greene and Cohen's moral revolution will occur and concludes that such a radical shift would precipitate practical difficulties not mentioned in their article. This Part, by contrast, sheds that assumption and asks whether society's intuitions will in fact change in ways that matter—that is, in ways that will trigger the problems associated with low moral credibility. This Part's discussion proceeds in three Sections. First, Section III.A examines a recent strand of scientific literature that both maps folk intuitions about criminal responsibility and explains what situational factors drive those intuitions. Second, Section III.B describes the new science of human agency and assesses whether it will reinforce or shape the intuitions charted in Section III.A. Finally, Section III.C builds this Note's central thesis by synthesizing these piecewise conclusions into a single model that predicts a forthcoming moral credibility problem.

Before delving into the science, though, we must specify which folk intuitions are relevant to examine and predict. Recall that one key aspect of Greene and Cohen's predictive account is that the new science of human agency will prove that both determinism and incompatibilism (or "hard determinism")

are accurate descriptions of reality.[129] Thus, in their view, every event in the natural world—including all human thought and behavior—is wholly the product of a complex matrix of causal antecedents and that state of affairs renders the man-made concepts of free will, blame, and responsibility nonsensical.[130] In their view, you do not deserve blame for reaching into the cookie jar; you were the hapless puppet of neuronal and environmental circumstances beyond your control. They further argue that society writ large will also grow to adopt that two-pronged worldview—one that both accepts determinism and rejects free will and responsibility.[131]

The only communal view relevant for present purposes, however, is a communal view that could clash with the principles of morality reflected in criminal law. And because the criminal law does not reflect or depend on a deterministic worldview, folk intuitions about free will and responsibility are all that matter. Whereas a determinist-incompatibilist from Greene and Cohen's future world would disagree with criminal law's current fixation on retributivism and be offended by a defendant's inability to claim a volitional impairment defense, a determinist-compatibilist would see no problem with either. Thus, the analysis below does not ask whether the new science of human agency will produce determinists, but rather how the new science of human agency will affect intuitions about free will and responsibility.

A. Folk Intuitions: Facts and Determinants

Although some scholars argue that a modern, scientific worldview presupposes determinism as an accurate description of reality,[132] determinism is less popular among laypeople.[133] Most people not only understand themselves as having robust free will,[134] but also downplay the extent to which deterministic factors like biology and environmental circumstances drive behavior.[135] In one

---

129 *See* Greene & Cohen, *supra* note 20, at 1776. Note that this position leaves open the possibility of some notion of responsibility, although not the sort of responsibility one might typically imagine. On their view, a criminal in a deterministic world is responsible for his actions insofar as he can be held accountable on consequentialist grounds. *Id.* at 1783.

130 *See id.* at 1780 (developing a hypothetical "Mr. Puppet" to examine the fallacy of human free will).

131 *Id.* at 1776.

132 *See* Hill, *supra* note 18, at 291, 330; Shariff et al., *supra* note 35, at 1563 ("Although few people deny that humans regularly make uncoerced choices and exercise self-control, many scientists and philosophers have taken issue with the idea that conscious humans can generate spontaneous choices and actions not fully determined by prior events.") (citation omitted).

133 *See* sources cited *supra* note 25 and accompanying text.

134 *See* Atiq, *supra* note 3, at 478; Shariff et al., *supra* note 35, at 1563.

135 *See* Atiq, *supra* note 3, at 486-87.

study, for example, participants were asked to indicate what percentage of a given criminal act stemmed from free will as opposed to hard social conditions.[136] On average, respondents attributed seventy-six percent of the conduct to free will and almost one third said the act was completely (100 percent) willed.[137] Further, societal beliefs in free will are similarly sturdy. Another study found that subjects clung to a belief in free will and moral responsibility even when asked to assume the truth of determinism.[138] Thus, for most people, free will is both real and possible in a deterministic world.[139]

In one sense, free will's durability and widespread appeal is not surprising. Johnson's aphorism rings true in that our everyday experiences seem to constantly confirm free will's presence. For example, the act of wagging a finger after commanding oneself to do so produces a powerful emotion of authorship. Intuitive appeal aside, there may be a better explanation for free will's popularity—namely, empirical evidence indicating that a belief in free will may itself be the product of entrenched cognitive processes operating beneath the level of conscious awareness.[140] Theorists have pointed to a number of different systemic features of our psychology in support of this idea,[141] among them the so-called fundamental attribution error. This error describes people's tendency to explain human behavior in dispositional rather than situational terms.[142] Because research shows that "individuals are especially prone to underemphasize the role of situational factors in the context of crime and punishment,"[143] one practical upshot might be that folk explanations for behavior prevail over scientific ones both in the jury room and as a general matter. Because reviewing the remainder of theories supporting the idea that free will and moral responsibility are to some degree structurally entrenched within our psyche would take more space than this Note can spare, suffice it to say that the idea is controversial, yet plausible.

Free will's deep-rooted appeal seems to spell trouble for Greene and Cohen, since beliefs in free will predict stronger tendencies toward retributivism.[144] If

---

136 *Id.* at 487.

137 *Id.*

138 Nahmias, *supra* note 25, at 215.

139 *See id.* at 215-16.

140 *See* Davies, *supra* note 11, at 166-69. The supreme irony of this conjecture is not lost on the author.

141 *See id.* (discussing the dual affect the theories of naïve realism and apparent mental causation have on human perception of authorship). *See generally* DANIEL M. WEGNER, THE ILLUSION OF CONSCIOUS WILL (2002) (constructing the theory of apparent mental causation, which generally posits that the experience of conscious will results from a cognitive system that both produces an emotion of authorship and reinforces that emotion by producing subsequent, conflated causal explanations for the conduct in question).

142 *See* Atiq, *supra* note 3, at 476-77.

143 *Id.* at 476.

144 Shariff & Vohs, *supra* note 35, at 78.

systemic cognitive mechanisms motivate us to preserve our beliefs in free will and responsibility, then our intuitions may in fact never collide with the compatibilist policies currently reflected in criminal law. Despite advancements in science, people might nonetheless continue to see criminals as morally responsible because our concepts of free will and blame are both intuitive and, as Greene and Cohen admit, entrenched in our psychology.[145]

Research on the popularity of libertarianism and the psychology of mental causation,[146] however, only tells half of the story. The remainder of research mapping out communal intuitions shows that, despite an enduring belief in free will, people exhibit a tendency to absolve criminals of blame if the biological and situational antecedents of the criminal's behavior are vivid enough.[147] This tendency is both consistent with the research already discussed in this Part and supported by additional evidence. The research above indicates that people do not instantly revert to incompatibilism when forced to contemplate a deterministic universe; they preserve a faith in free will and responsibility. The set of studies described below shows that, as people learn about greater and greater portions of the causal matrix underlying different thoughts and behaviors, they become incrementally less likely to apportion blame regardless of their beliefs in free will.

Findings drawn from one of the first[148] large-scale attempts to measure society's moral intuitions support the related premise that attributions of blame diminish stepwise alongside the degree of perceived control the criminal actor has. Professors Paul Robinson and John Darley uncovered this pattern while performing research explicitly designed to compare communal views with the letter of the law.[149] Their research design was simple. Subjects read a series of scenarios depicting various crimes being committed and assigned a liability score to the criminal in each scenario.[150] The liability scores subjects gave imaginary defendants in classic excuse defense scenarios—including insanity, involuntary intoxication, and duress—are particularly illustrative of the sliding scale of liability and control described above.

---

145 Greene & Cohen, *supra* note 20, at 1782 (reporting research showing that "humans have a set of cognitive subsystems that are specialized for processing information about intentional agents").

146 The term "mental causation" refers, as a general matter, to the mind's tendency to drive the subjective experience of willed agency even if that conclusion departs from reality. *See* sources cited *supra* note 141.

147 *See* ROBINSON & DARLEY, *supra* note 33, at 127-50; Nahmias, *supra* note 25, at 230-31; Shariff et al., *supra* note 35, at 1568; Shariff & Vohs, *supra* note 35, at 78.

148 ROBINSON & DARLEY, *supra* note 33, at 2-3 (asserting that, as of 1995, neither social scientists nor psychologists had "mapped the contours of the moral intuitions of our culture").

149 *Id.* at 3, 130-50.

150 *Id.* at 7.

To test whether society's intuitions validate or call into question the many different forms of the insanity defense, Robinson and Darley devised seven scenarios in which an imaginary culprit exhibited varying degrees of cognitive deficiency (inability to know right from wrong) and control deficiency (inability to avoid committing the crime).[151] Results indicated that although both forms of deficiency justified lower liability scores, control dysfunction elicited a more powerful mitigation effect.[152] Further, the greater the degree of control dysfunction, the lower the liability scores.

Subjects produced similar scores in the context of both involuntary intoxication and duress. Researchers devised involuntary intoxication scenarios that also varied by degree of control and cognitive dysfunction because, for most states, elements of the involuntary intoxication defense mimic those of the insanity defense.[153] Although subjects thought that involuntarily intoxicated defendants were more blameworthy than insane defendants, high levels of dysfunction predicted proportionally low liability scores in both cases.[154] The duress scenarios depicted defendants acting under varying degrees of coercion (from no coercion to a threat to murder the defendant's entire family) instead of dysfunction, but, predictably, also elicited liability scores arranged on a continuum that depended on the defendant's level of control.[155]

In addition to showing that society sees degrees of liability instead of criminal law's absolutes of guilty and not guilty, Robinson and Darley's findings provided the first hints that society equates past causes with exonerating excuses. Each cause, no matter its type—biological, biochemical, situational, or otherwise—mitigated subjects' blame attributions to a degree commensurate with the cause's potency.

Recent studies have added a finer point to this sliding scale model.[156] They reveal that learning about the low-level causal mechanisms underlying human behavior—both with respect to a token actor's conduct[157] or generally applicable to all behaviors[158]—reduces retributivism. Importantly, this correlation holds regardless of the subjects' beliefs in free will.[159] Thus, knowledge of or exposure to deterministic explanations of behavior appears to independently reduce people's willingness to dole out just deserts.

---

151 *Id.* at 130-33; *see also id.* app. A at 262-65 (providing the full text of the scenarios).
152 *Id.* at 134.
153 *Id.* at 139-40.
154 *Id.* at 155.
155 *Id.* at 147-50.
156 *See* Shariff et al., *supra* note 35.
157 *See* Nahmias, *supra* note 25, at 230.
158 *See* Shariff et al., *supra* note 35, at 1568; Shariff & Vohs, *supra* note 35, at 78.
159 Nahmias, *supra* note 25, at 230.

A study conducted in 2006 revealed that simply exposing people to deterministic explanations for behavior does not impact their moral intuitions.[160] Instead, what matters is explaining those behaviors using precise, reductionistic terms.[161] Subjects were given two descriptions of a deterministic world and asked whether actors in each world deserved blame for their actions. However, descriptions of each world differed as follows: human behaviors in the reductionistic world were described as "*completely caused* by the particular chemical reactions and neurological processes occurring in their brain,"[162] whereas human behaviors in the nonreductionistic world were described as "*completely caused* by the particular thoughts, desires, and plans they have."[163] Responses diverged widely. Of those given the reductionistic description, only eighteen percent and nineteen percent indicated free will and moral responsibility were possible, respectively.[164] Those given the nonreductionistic description, by contrast, exhibited much greater ontological optimism—seventy-two percent allowed for free will and seventy-seven percent for moral responsibility.[165]

One interpretation of these results is that libertarian intuitions remain strong in the face of determinism unless the deterministic picture relayed is causally comprehensive or otherwise vivid. Thus, its results are consistent with those of Robinson and Darley because both suggest that, as science crowds folk psychological explanations of behavior out of a causal chain, people become gradually less retributive. This study suffers from major limitations, however, including a small, nondiverse sample size (forty-nine college students),[166] short and simple explanations of both the reductionistic and nonreductionistic worlds, and an experimental design that in no way resembles real-world blame attributions.

A set of studies published in 2014 addressed each of these methodological flaws and reached similar results.[167] Researchers first exposed subjects to deterministic concepts through one of three means: a scholarly article arguing against free will; a popular science magazine article describing mechanistic neural processes but not mentioning free will; and a semester-long introductory neuroscience course.[168] After exposure, subjects answered a questionnaire about a fictional murderer designed to measure their desire for retributive punishment.[169]

---

160 *Id.*
161 *See id.*
162 *Id.* at 230-31 (emphasis in original).
163 *Id.* at 231 (emphasis in original).
164 *Id.*
165 *Id.* at 231, 233.
166 *Id.* at 230 n.16.
167 *See* Shariff et al., *supra* note 35, at 1564.
168 *Id.* at 1565-68.
169 *Id.* at 1565-66.

Each experiment rendered the same result: exposure to sophisticated, fine-grained explanations of the neural substrates underlying human behavior reduced retributivism.[170]

Two important limitations of the 2014 studies bear mentioning. First, they do not indicate how long the effects of exposure to deterministic explanations lasts. A reduction in retributivism that lasts only the few hours during which the vivid science is top-of-mind has little to no practical import for present purposes. In order to meaningfully impact moral credibility, a given individual's shift in moral values must last until exposure to the relevant criminal law. Second, the various methods of exposure to the science were atypical. Not many people take neuroscience courses, read science magazines, or read groundbreaking scholarly theses. The more likely method of exposure—a popular news article—may fail to provide the same level of granularity. The 2006 study addresses this limitation to some degree, however, because it changed subjects' intuitions despite only describing a marginally complex description of reality.[171]

Empirical indicators of society's moral propensities are new, but nonetheless reveal a number of patterns relevant to determining whether the new science of human agency will change society's views in ways that undermine retributivism's moral credibility. First, although widespread beliefs in free will are sturdy, people's views on moral responsibility appear to readily change upon exposure to scientific explanations for behavior that crowd out the possibility of human agency. Second, there is some evidence to suggest that this dynamic operates on a sliding scale: the clearer the scientific explanation, the greater the effect on subjective blame attributions. Finally, despite the limitation noted above, the functional triggers of these moral shifts can be commonplace—reading a magazine will do the trick. Thus, the science reasonably suggests that when the following three criteria obtain, a given actor's tendency for retributivism will diminish: (1) exposure (2) to a reductionistic, granular explanation of human behavior (3) that crowds out any explanations based on folk psychological mental states.[172]

Here, the term "folk psychological mental states" refers to colloquial concepts used to describe behavior that are readily definable and coherent in everyday discourse, but opaque in strictly scientific and materialistic terms. For example, explaining that a burglary happened because the thief "was an unscrupulous guy who wanted to get rich" is a folk psychological explanation.

---

170 *Id.* at 1568.

171 *See* Nahmias, *supra* note 25, at 230-31 (describing two relatively simple conceptions of free will that were presented to research subjects).

172 Eddy Nahmias describes "folk psychology" as "inherently non-reductionistic, explicitly requiring a role for conscious beliefs, desires, reasons, plans, and deliberations to cause our choices and actions." *Id.* at 229.

Unlike reductionistic explanations, folk psychological ones invite blame because they seem to presuppose libertarian free will. The research supports this intuition; if people contemplate a deterministic universe described using folk psychological mental states instead of scientific, reductionistic terms, those people are more likely to preserve free will and moral responsibility.

The science reviewed in this Section is by no means a comprehensive review of the evidence relevant to addressing the substantiation objection. The volume and breadth of evidence and causal interactions one would need to examine to confirm or deny Greene and Cohen's prediction on empirical grounds would be breathtakingly vast. Nonetheless, the modest body of work reviewed in this Section suggests that, at this juncture, the idea that advancements in neuroscience and related fields can change society's moral intuitions regardless of whether it changes their views on free will is at least plausible. The next Section briefly scans the new science of human agency to evaluate the likely extent of that shift.

B. The New Science of Human Agency

In the Introduction, this Note defined the new science of agency as an emerging class of empirical literature characterized by a tendency to identify the specific neurological, genealogical, and endocrinal processes that underlie given behaviors and instances of conscious decisionmaking. Research on the causes and effects of moral evolutions, however, indicate that these new sciences have practical import beyond merely describing the natural world. To wit, the vivid causal pictures they paint may change traditional communal views on blame and responsibility in ways that may diminish retributive justice's moral credibility. This Section aims to identify the characteristics of the new science of human agency that make it especially likely to contribute a shift in our moral topography. It accomplishes that task by contrasting prior scientific attempts to explain behavior with more recent attempts.

Recall the criteria identified in Section III.A that predict a given actor's diminished sense of retributivism: (1) exposure (2) to a reductionistic, granular explanation of human behavior (3) that crowds out any explanations based on folk psychological mental states. Older studies positing deterministic explanations for human behavior generally flunk criteria (2) and (3) for a number of reasons.

First, they were largely correlational and failed to explain the low-level causal mechanisms that precipitated higher-order criminal behaviors.[173] For example, studies conducted before 1988 reported correlations between chromosomal abnormalities and increased aggression,[174] irregular

173 *See* Denno, *supra* note 9, at 619-40.
174 *Id.* at 620.

electroencephalograph readings and murderous tendencies,[175] and testosterone production and irritability,[176] but failed to explain the processes driving those correlations.[177] Thus, although these pre-1988 studies explained criminal behavior using the same sort of genealogical, neurological, and endocrinal evidence as the new sciences, their lack of specificity preserved large gaps in causal stories that laypeople could have filled with folk psychology.

Second, pre-1988 studies in fact invited folk psychological conjectures by relying on mental states, not physical mechanisms, as descriptive terms.[178] Studies reported genetic, hormonal, and cognitive abnormalities as predictive of "exaggerated maleness,"[179] "aggression,"[180] and "lack of discipline."[181] By using descriptors that are, to some degree, nonspecific and hard to pin down in materialistic terms, early empirics gave lay observers room to infer some degree of control on behalf of the criminal actor.

Finally, these correlational studies were widely prone to conflicting results.[182] If the scientific community was not convinced of the proposed causal mechanisms' explanatory potency, how could the public?

Next-generation research, by contrast, supplants correlational data with causal theories that delve deeply into the relevant physiological substrates of behavior.[183] Instead of descriptive terms that evoke mental causation, the new sciences refer to brain states using mechanistic terms like "diminished capacity"[184] and "organic impairment."[185] The human agent is thus better understood under these postulates as a biological machine instead of a thinking, planning actor. Further, because this deterministic conception of the human self "exerts a strong pull on the popular, educated imagination,"[186] mainstream media

---

175 Id. at 637-38.

176 Id. at 628.

177 *See* sources cited *supra* notes 174-176.

178 *See, e.g.*, Denno, *supra* note 9, at 620, 640.

179 *Id.* at 620.

180 *Id.* at 626.

181 *Id.* at 644.

182 *Id.* at 627-31, 634-36, 640, 646-48.

183 *See, e.g.*, Atiq, *supra* note 3, at 456-57; Daniel A. Martell, *Causal Relation Between Brain Damage and Homicide: The Prosecution*, 1 SEMINARS CLINICAL NEUROPSYCHIATRY 184 (1996) (hypothesizing that physiological abnormalities in a particular defendant's frontal lobe affected his volitional capacities); Peggy Sasso, *Implementing the Death Penalty: The Moral Implications of Recent Advances in Neuropsychology*, 29 CARDOZO L. REV. 765, 790-91 (2007) (discussing the behavioral effects of damage to the orbitofrontal cortex); Eagleman, *supra* note 5 (summarizing a body of research that, according to Eagleman, "demonstrates the limits of the [free choice] assumption").

184 Sasso, *supra* note 183, at 790.

185 Sapolsky, *supra* note 10, at 1794.

186 Morse, *supra* note 8, at 127.

outlets have disseminated stories that implicitly or explicitly endorse deterministic sciences at a relatively high clip.[187] For all of these reasons, the new science of human agency meets each of the three criteria of an intuition-mover outlined above.

Two examples best illustrate this new breed of research. Consider first the example of the gene for monoamine oxidase A (MAOA).[188] Statistical data evidences a robust correlation between the presence of a lower-activity MAOA variant and criminally deviant behavior, but scientists have only recently discovered why.[189] If an MAOA carrier experiences certain environmental stimuli, this MAOA variant triggers a neurochemical response system that leads to a "functional difference" in his brain regions responsible for "anger production and control."[190] Note the temporal cohesiveness of this causal picture; it describes the mechanisms responsible for the potentiality, genesis, and actualization of antisocial behavior. Although presence of the MAOA variant is by no means an independent predictor of criminality, this theory provides the sort of vivid causal story that may diminish subjective attributions of blame towards carriers of the lower-activity MAOA variant.

A second and perhaps more powerful example of the new science of human agency is Sapolsky's account of prefrontal cortex (PFC) damage. The PFC has a number of important functions, but most relevant for present purposes is its job of "biasing an individual towards doing the 'harder' but 'more correct' behavior" instead of impulsively succumbing to the choice that provides instant gratification.[191] Unfortunately for us, the PFC is prone to underdeveloping, sustaining damage, or otherwise misfiring for any number of reasons—namely, age, transient states of intoxication, blunt trauma, lesion, tumor, and neuronal

---

187 *See, e.g.*, sources cited *supra* note 5. Perhaps the most notable example of this sort of widespread media coverage has been the story of Jim Fallon, a neuroscientist who accidentally realized his brain shared the same physiological characteristics as murderous psychopaths. Since his discovery, Fallon has featured in news stories, spoke in two TED talks, and wrote a *New York Times* bestselling book called *The Psychopath Inside*. Fallon, *supra* note 36; Susan Donaldson James, *Scientist Related to Killers Learns He Has a Psychopath's Brain*, ABC NEWS (Nov. 30, 2013), http://abcnews.go.com/Health/scientist-related-killers-learns-psychopaths-brain/story?id=21029246; Fallon, *supra* note 6.

188 Atiq, *supra* note 3, at 456.

189 *Id.* at 456-57; Matthew L. Baum, *The Monoamine Oxidase A (MAOA) Genetic Predisposition to Impulsive Violence: Is It Relevant to Criminal Trials?*, 6 NEUROETHICS 287, 288 (2013) ("[C]arrying the low activity MAOA gene (MAOA-L) could make the subject more prone to express aggression if provoked or socially excluded."). An Italian appeals court in 2009 reduced a defendant's sentence for murder based on evidence that he carried this genetic variant. *Id.* at 287.

190 *Id.* at 456 (naming "childhood maltreatment" as an example of an environmental stimulus).

191 Sapolsky, *supra* note 10, at 1793.

death.[192] When this happens, concomitant behavioral changes are remarkable. Studies have shown that, when presented with two options, subjects with damaged PFCs can verbalize an intent to choose the "right" option but then prove literally unable to avoid choosing the wrong one.[193] Further, this effect is proportional.[194] Thus, the net functional capacity of any person's PFC, measurable by metabolic rate, depicts that person's capacity to regulate impulsivity.[195] To summarize: your PFC is one yardstick (surely of many) of your volitional capacities.

Admittedly, research on the MAOA variant and PFC damage were chosen because they are especially compelling illustrations of how advancements in neuroscience and related fields are challenging traditional concepts of the human self. But that is precisely the point. Studies like these leave little space for folk psychology to infiltrate the causal matrix that governs behavior; in essence, they make determinism somewhat intuitive. For that reason, it is plausible to conclude that, in light of the three criteria above, these and related discoveries will begin to change our moral intuitions if widely disseminated. This is especially true given that, by all accounts, even Stephen Morse's,[196] scientific advancements in these fields are occurring at a blistering pace.[197]

C. Bringing It All Together: Two Dimmers and a Flip Switch

This Part has sought to contribute to the debate on whether scientific advancements justify reforming criminal law by answering a core empirical question: whether the new science of human agency will change society's views on free will and responsibility, and if so, in what ways? Answering this question accomplishes at least two things: it both addresses the substantiation objection and illuminates another argument that reformists might make—that the new science of human agency will have the practical effect of eroding retributivism's moral credibility.

The research reviewed in Sections III.A and III.B indicates that exposure to the new science of human agency can change individuals' views on responsibility, but not necessarily free will. While beliefs in free will appear to remain sturdy in the face of deterministic explanations for human behavior, those same granular explanations diminish individuals' willingness to apportion blame

---

192 *Id.*

193 *Id.*

194 *Id.*

195 *Id.* at 1793-94.

196 *See* Morse, *supra* note 8, at 123 (indicating that neuroscientific evidence may soon develop to the point at which it is regularly admitted in trial proceedings).

197 *See* Hill, *supra* note 18, at 291; Justickis, *supra* note 41, at 233 (referring to psychological advancements); Shariff & Vohs, *supra* note 32, at 78.

for given criminal acts. Unlike beliefs in free will, blame attributions are fluid and responsive to the presence or absence of ascertainable causal antecedents that crowd out folk psychological explanations for behavior. Further, retributivism appears to diminish stepwise as the perceived causal power of scientific explanations increases. The new science of human agency is in the business of illuminating a growing spectrum of unconscious and uncontrolled determinants of human behavior. It is therefore plausible to conclude not only that individuals' propensities to punish retributively will decrease upon sufficient exposure to the relevant science, but also that this effect will intensify over time.

These empirical findings lend little support to Greene and Cohen's prediction that the new science will precipitate a widespread rejection of free will, moral responsibility, and aspects of our criminal justice system that seem to depend on those two concepts.[198] Instead, the science reviewed by this Note supports the comparatively modest empirical premise that exposure to deterministic explanations of human behavior makes people view particular criminals as somewhat less responsible for token crimes. Thus, although intuitions might shift, this Note found no empirical evidence that those intuitions will shift in the ways Greene and Cohen predict. Most people believe that free will and blame apply to causally determined behaviors, albeit to varying degrees. Whether they know it or not, most laypeople are compatibilists.

Although the body of research reviewed above does not alleviate the substantiation objection, it does present a new argument for reformists—namely, that changes to the criminal law may be necessary to avoid a moral credibility problem. Recall from Part II that some evidence indicates criminal law suffers from diminished efficacy when it produces practical outcomes that people deem unjust. If, in the context of specific criminal cases, reductionistic explanations for the deviant behavior exist but nonetheless fail to ensure acquittal or lesser sentences, moral mismatches might occur once people experience direct or indirect exposure.

To clarify this argument, consider the fact that the empirical strands summarized in Parts II and III feature proportional continua: (1) the criminal law's efficacy diminishes gradually alongside the widening gap between people's moral intuitions and legal policies; and (2) defendants are seen as proportionally less culpable as the physiological correlates of their behavior come clearer into the fore. These can be seen as dimmer switches. The criminal law, however, is better understood as a flip switch that primarily deals in bimodal absolutes—guilty and not guilty.

Although sentencing guidelines enable judges and juries to apportion punishment based on a given allowable range, criminal law is nonetheless not as

198 *See* Greene & Cohen, *supra* note 20, at 1776.

flexible as our two dimmer switches for two reasons. First, exposure vehicles—namely television and newspaper stories—might focus primarily on the presence or absence of conviction and gloss over the significance of the sentence handed down. In this scenario, those exposed to the broadcast would necessarily form a moral judgment based on incomplete information: a verdict and a sentence reported absent an explanation of the sentence's severity as a relative matter. Thus, moral mismatches might form based primarily on verdicts because lay listeners, for whatever reason, fail to appreciate the significance of sentences. Second, sentencing guidelines have procedural constraints that bound the continuum. These two descriptors substantiate the intuitive premise that the law's rule-driven processes lack the seemingly limitless malleability of individuals' shifting conceptions of justice.

To extend the metaphor, as the new science of human agency gradually slides our retributivism dimmer switch towards "off," the criminal law flip switch will fail to capture the same nuanced, stepwise decrease in the punishments it hands down. Resultantly, our moral credibility dimmer switch will slide towards "low" to a degree proportional to the mismatch created above. This rudimentary model illustrates the empirical reality that as the gap between retributivist tendencies and criminal law outcomes widens, moral credibility diminishes. And a broad range of recent empirics suggests that the new science of human agency is particularly likely to set this corrosive sequence in motion. Thus, unless the criminal law adapts to accommodate our developing understanding of human decision making processes, our justice system might soon face a moral credibility problem.

CONCLUSION

Retribution—the idea that we ought to punish criminals simply because they deserve it—plays a critical role in American legal discourse and policy. Indeed, the degree to which a defendant is morally culpable formally justifies, in part, imposition of the U.S. criminal justice system's ultimate sanction: the death penalty.[199] This Note's findings provide one practical reason why scientific developments should inform the manner in which we apply the concepts of responsibility and blame in given scenarios—namely, that failure to do so will diminish criminal law's efficacy.

That is not to say, however, that our justice system has demonstrated a complete failure to adapt in light of available empirics. In *Roper v. Simmons*, for example, the U.S. Supreme Court cited research on developmental psychology to

199 Roper v. Simmons, 543 U.S. 551, 571 (2005) ("We have held there are two distinct social purposes served by the death penalty: 'retribution and deterrence of capital crimes by prospective offenders.'") (quoting Atkins v. Virginia, 536 U.S. 304, 319 (2002)).

support its holding that executing a minor violates the Eighth Amendment's ban on cruel and unusual punishment.[200] There, the Court reasoned that a minor's irresponsible conduct is "not as morally reprehensible as that of an adult" due to observable differences in minors' decision-making and volitional capacities.[201] To wit, "scientific and sociological studies" confirmed that youths are categorically more impulsive, susceptible to environmental pressures, and affectively malleable than adults.[202] In light of that available evidence, the Court explained, youths are less blameworthy and therefore undeserving of extreme retribution.[203] The *Roper* Court thus explicitly attempted to accomplish that which the science reviewed in this Note recommends: reducing criminal liability to a degree commensurate with a given criminal actor's organic volitional capacities.

In addition to supporting the relatively uncontroversial idea that the law should periodically adapt in light of changing scientific realities, this Note's findings inform a far more contentious debate: whether mounting evidence that human behavior stems from biological, mechanistic causes suggests that our criminal law should dispense with retributivism altogether. It does so by using empirical evidence to construct the following argumentative framework upon which both empiricists and legal theorists might build. As the new science of human agency gradually illuminates an increasing number of links in the causal chain underlying given criminal behaviors, laypeople will view those behaviors as less blameworthy; and because the criminal law currently offers defendants limited means to avoid liability on the basis of diminished volitional capacity, laypeople's moral views will increasingly clash with criminal law outcomes. These moral mismatches will, in turn, precipitate diminished compliance, cooperation, and acquiescence with criminal laws, as well as increased general deviance. In short, available empirics suggest that changing public perceptions will bring adverse practical effects if the law remains static.

This argument—which amounts to an empirical prediction—supports reformists' charge, but cannot justify broad-based reforms to our justice system per se for two reasons. First, the looming threat of diminished moral credibility is one of many factors surely relevant to deciding whether retributivism's costs outweigh its benefits. Indeed, the age-old debate surrounding proper

---

200 *Id.* at 569-70, 573 (relying on "scientific and sociological studies" as well as diagnostic practice in psychiatry).

201 *Id.* at 570 (quoting Thompson v. Oklahoma, 487 U.S. 815, 835 (1988)).

202 *Id.* at 569-70.

203 *See id.* at 570. ("[T]he case for retribution is not as strong with a minor as with an adult. Retribution is not proportional if the law's most severe penalty is imposed on one whose culpability is diminished, to a substantial degree, by reason of youth and immaturity.").

justifications for criminal punishment spans centuries, oceans, and disciplines,[204] and cannot be resolved by a single practical consideration. Thus, this Note's argument is best considered a single weight on the scale in favor of reform, not a debate-settling silver bullet. Second, methodological limitations in the research constrain the generalizability of present findings. The research does not tell us, for example, how frequently people are exposed to the new science of human agency, how long exposure to the new science of human agency affects individuals' moral judgments, or the extent to which moral mismatches drive general criminal deviance. Thus, the degree of the looming moral credibility problem is currently unknown. It would be premature to reorganize our criminal justice system based on data that is, in some sense, incomplete.

Nonetheless, the thesis of this Note exposes our criminal law's dependence on blame and responsibility as potentially antithetical to its own animating purposes. As such, it naturally invites speculation and conjecture on an enticing question this Note has, for reasons already discussed, hesitated to confront: what would a criminal justice system without the concepts of retribution, blame, and responsibility look like? In the last few lines before concluding, I succumb to temptation and offer a brief, 50,000-foot suggestion informed by the findings recounted above.

The broad-strokes solution is simply stated: craft a system whose sole focus is to serve the utilitarian goals of deterrence, rehabilitation, and incapacitation. As noted in Section I.A, a criminal convicted in this system—blameworthy or not—would receive a sentence designed to optimally benefit society. Importantly, however, this system would not necessarily dispense with retribution as an articulated goal of criminal punishment or remove the concepts of blame and responsibility "from the legal argot."[205] Instead, because available empirics reveal that human perception is one guarantor of (or cancer to) the system's effectiveness, this system would preserve those concepts to the extent that they serve, among other utilitarian ends, the function of producing outcomes in line with then-existing moral views. For example, a hypothetical jury explicitly instructed to consider desert and blameworthiness might hand down a verdict that better harmonizes with widely held moral beliefs.[206] Similarly, this system would avoid handing down outlandishly severe punishments for comparatively minor offenses, even if doing so were considered a maximally

---

204 *See, e.g.*, Matthew Haist, Comment, *Deterrence in a Sea of 'Just Deserts': Are Utilitarian Goals Achievable in a World of 'Limiting Retributivism'?*, 99 J. CRIM. L. & CRIMINOLOGY 789, 790-92 (2009) (providing a brief description of this debate's history, scope, and significance).

205 Eagleman, *supra* note 5.

206 In some jurisdictions, judges can and do instruct jurors to consider the purposes of punishment in making their decision. *See* Cotton, *supra* note 26, at 1317.

effective deterrent and rehabilitator, if that course of action sowed widespread discontent and a corresponding diminishment of the system's moral credibility that offset originally intended benefits. Thus, the concepts of retribution, blame, and responsibility would function as tools readily manipulated in service of utility, not, as is currently the case, bedrock precepts that limit criminal law's responsiveness to ever-changing moral tides.

This model is crude and oversimplified, but nonetheless addresses the core problem this Note has identified and offers the beginnings of a solution upon which others may build. Further, it is in some sense noncontroversial; reformists and conservationists agree that our system's utilitarian goals are worth preserving. Retribution, by contrast, has proved a key sticking point due to that concept's uncertain future. Thus, a model that repurposes retribution to serve utilitarian ends strikes a compromise between those unsure of the normative implications of the new science of human agency and those who champion that science as the harbinger of retributive justice's demise. Perhaps less importantly, it excavates the fate-versus-free will debate from its dusty, ancient seat of practical irrelevance.