2019

Neuroscience, Justice and the "Mental Causation" Fallacy

John A. Humbach
Elisabeth Haub School of Law at Pace University

Washington University Jurisprudence Review

VOLUME 11 NUMBER 2 2019

NEUROSCIENCE, JUSTICE AND THE "MENTAL CAUSATION" FALLACY

JOHN A. HUMBACH*

> All actions take place in time by the interweaving of the forces of Nature; but the man lost in selfish delusion thinks that he himself is the actor.[1]

ABSTRACT

Mental causation is a foundational assumption of modern criminal justice. The law takes it for granted that wrongdoers "deserve" punishment because their acts are caused by intentions, reasons and other mental states. A growing body of neuroscience evidence shows, however, that human behavior is produced by observable physiological activity in the brain and central nervous system—all in accordance with ordinary physical laws. Beyond these ordinary physiological interactions and processes, no hypothesis of mental causation is required to causally explain behavior.

Despite the evidence, neuroskeptics insist that intentions, reasons and other mental states can play a causal role in producing human behavior. The evidentiary case for mental causation turns out, however, to be premised on a well-known logical fallacy, post hoc ergo propter hoc.

* Professor of Law, Elisabeth Haub School of Law at Pace University.

Note to Readers: It should generally not be necessary to read the footnotes, sometimes lengthy, in order to follow the main points of this article. They are meant mainly as supplementation and as a place to comment on technical issues that may be of interest to some but a distraction from the main flow for others.

1. BHAGAVAD GITA 3:27–28 (Juan Mascaró, trans. 1962) (c. 500 B.C.).

Meanwhile, based on the best explanation of all the evidence and data, mental causation almost certainly cannot and does not occur.

If mental causation is the basis on which offenders are deemed to deserve punishment, current punishment practices may need to be revised in the interest of justice. While society will probably always need to use coercive measures against persons who pose intolerable dangers and risks, the nature and quality of those measures may be very different if they are treated as a regrettable necessities rather than as deserved.

INTRODUCTION

The idea that mental states such as intentions and reasons can cause or influence human behavior is thickly woven into criminal law. The multiplicity and substantive importance of legally relevant mental states (intention, purpose, volition, etc.), each having its own particular significance in assessing guilt,[2] leaves no doubt that modern conceptions of justice are deeply dependent on the "folk psychology" assumption that wrongful acts are the products of culpable minds.[3] In sharp distinction to most of the law's assumptions about the world, however, the putative causal efficacy of mental states does not even purport to be based on known facts about physical reality.

In recent decades, a growing body of neuroscience research has developed an alternative causal explanation of how human behavior is produced.[4] According to this newer explanation, the things that people do are caused by their brains, not by their minds or mental states.[5] The newer

---

2. *See generally* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 117–44 (6th ed. 2012); Wayne R. LaFave, CRIMINAL LAW 252–88 (5th ed. 2010),

3. *See* Stephen J. Morse, *The Inevitable Mind in the Age of Neuroscience* 34, *in* PHILOSOPHICAL FOUNDATIONS OF LAW AND NEUROSCIENCE (Patterson et al. eds., 2016) [hereinafter Morse, *Inevitable Mind*]; Stephen J. Morse, *Determinism and the Death of Folk Psychology: Two Challenges To Responsibility from Neuroscience*, 9 MINN. J.L. SCI. & TECH. 1, 2–3, 10–11 (2008) [hereinafter Morse, *Folk Psychology*] ("Roughly speaking, the law implicitly adopts the folk-psychological model of the person, which explains behavior in terms of desires, beliefs and intentions."); Stephen J. Morse, *Lost in Translation: An Essay on Law and Neuroscience* 530 (2011), *in* LAW AND NEUROSCIENCE (Michael Freedman, ed. 2011) [hereinafter Morse, *Translation*].

4. A study conducted by Elsevier found that 1.79 million articles were published in the area of brain and neuroscience research during the period 2009 to 2013. Georgin Lau et al., *New Report Maps The Landscape Of Global Brain Research* (2014), https://www.elsevier.com/connect/new-report-maps-the-landscape-of-global-brain-research; *see also* RICHARD PASSINGHAM, COGNITIVE NEUROSCIENCE: A VERY SHORT INTRODUCTION 3 (2016) ("[N]early 30,000 experiments conducted using fMRI alone."). Much of this research is described and summarized in ROBERT M. SAPOLSKY, BEHAVE: THE BIOLOGY OF HUMANS AT OUR BEST AND WORST (2017).

5. Just to be clear, it is not meant to suggest any event ever has a single cause; rather, every caused event has multiple interacting causes running back indefinitely in time. Statements that "the brain causes behavior" or the like should be understood as a shorthand for saying that the brain and

causal explanation presupposes that the brain produces bodily movements entirely by means of ordinary physical forces and in accordance with ordinary physical principles.[6] By contrast, the mental-cause hypothesis seems to suppose the existence of forces that have no counterparts anywhere else in the physical domain. The supposed causative power of intentions, reasons and other mental states is conceived instead to be a unique and extraordinary property of minds (whose existence is, of course, also inexplicable in its own right). The mental-cause explanation of human behavior bears, in other words, the distinctive mark of an *ad hoc* gap filler, a made-to-order tale of cause and effect that has been devised, for perhaps ulterior reasons of policy,[7] to account for an otherwise mysterious correlation—the apparent link between conscious thoughts and subsequent conduct.

The question addressed in this article is whether the growing body of neuroscience evidence should make a legal difference.[8] Many are openly skeptical and argue that it should not.[9] This skepticism no doubt results in

---

central nervous system—and, indeed, the body's physiology generally—play a decisive *causal role* in directing and producing an individual's behavior. That is, "brain causation" refers metonymically to all of the biomechanical causes of behavior that operate "inside the skin," as it were, with the exception of possible mental-state causes.

6. My understanding of the neuron-based description of human behavior is based primarily on IRA B, BLACK, INFORMATION IN THE BRAIN (1991) (focusing on molecular level); JEAN-PIERRE CHANGEUX, NEURONAL MAN: THE BIOLOGY OF THE MIND (1985) (general introduction to brain structure and its functioning in information processing); PATRICIA S. CHURCHLAND, NEUROPHILOSOPHY: TOWARD A UNIFIED SCIENCE OF THE MIND/BRAIN (1986) (comprehensive essay relating neurophysiological findings to the perennial "mind/body" problem); PATRICIA S. CHURCHLAND & TERRENCE J. SEJNOWSKI, THE COMPUTATIONAL BRAIN (1992) (information processing across biological neural networks); ANTHONY SAMASIO, DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN (2005); DAVID H. HUBEL, EYE, BRAIN AND VISION (1988) (brain information processing with emphasis on visual information); BRYAN KOLB & IAN Q. WINSHAW, FUNDAMENTALS OF HUMAN BRAIN AND BEHAVIOR (4th ed. 2012); PASSINGHAM, *supra* note 4, at 73; ROBERT M. SAPOLSKY, BEHAVE: THE BIOLOGY OF HUMANS AT OUR BEST AND WORST (2017); *see also* DANIEL J. AMIT, MODELING BRAIN FUNCTION (1989) (introducing a mathematical model of brain decision function); ARNOLD TREHUB, THE COGNITIVE BRAIN (1991) (a neurophysiological account of human cognitive processing); DANIAL DENNETT, CONSCIOUSNESS EXPLAINED 162–66 (1991) (a very readable tour-de-force on the mind as the work of the brain); PAUL M. CHURCHLAND, A NEUROCOMPUTATIONAL PERSPECTIVE (1989).

7. *See infra* note 24; Morse, *Inevitable Mind*, *supra* note 3, at 46–47; Morse, *Translation, supra* note 3, at 543 (quoted infra note 24); Morse, *Folk Psychology*, supra note 3, at 2–3.

8. This is not a topic of passing interest: the number of neurolaw publications has grown from around 100 to over 1600 in just the past 10 or so years. Owen D. Jones & Anthony D. Wagner, *Law and Neuroscience: Promise, Progress and Pitfalls*, *in* THE COGNITIVE NEUROSCIENCES 3 (Michael Gazzaniga et al. eds., 2019). Among notable examples are Joshua Greene & Jonathon Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, PHIL. TRANS. ROYAL SOCIETY 1775–1785 (2004); LAW AND THE BRAIN 207 (Semir Zeki & Oliver R. Goodenough eds., 2006); Robert M. Sapolsky, *The Frontal Cortex and the Criminal Justice System*, 359 PHIL. TRANS. R. SOC. LOND. 1787 (2004) (stressing the persisting "two cultures" problem that divides legal and scientific modes of thinking and hinders communication of advances in the latter field for incorporation into the former).

9. *See, e.g.*, *infra* note 47 ("[N]euroscience has little to contribute"); Morse, *Translation*, *supra* note 3, at 534 ("The law will be fundamentally challenged only if neuroscience or any other science

part from the fact that the neuroscience view of behavior causation is at odds with the law's foundational assumption that offenders "deserve" to be punished because their actions are due to culpable mental states.[10] This neuroskeptical position will be examined and found wanting.

In Part I, we review the central role of mental causation in our thinking about criminal justice. In Part II, the article looks more closely at the core of the skeptical view of neuroscience. Part III considers the threat that the findings of neuroscience pose to the traditional mental-causation justifications for punishment. In Part IV, there will be a review of the meager evidence that mental causation occurs and of the logical fallacy that is commonly employed in drawing inferences from that evidence. Part V is a comparative evaluation of the case for mental causation and its newer rival, the causal explanation provided by neuroscience, using the methodology of "inference to best explanation." Finally, in Part VI, the article will close with a consideration of some of the implications for criminal justice if the mental-causation hypothesis is supplanted by the biomechanical alternative that is provided by neuroscience.

---

can conclusively demonstrate that the law's psychology is wrong and we are not the type of creatures for whom mental states are causally effective"). A complete list of neuroskeptical writings would be impossible here, but some that will be mentioned further on include: Steven K. Erickson, *Blaming the Brain*, 11 MINN. J L. SCI. & TECH. 27 2010; Iskra Fileva & Jonathon Tresan, *Will Retributivism Die and Will Neuroscience Kill It?* COGNITIVE SYSTEMS RESEARCH 34–35 (2015) (concluding that neuroscience does not disprove free will—not focusing on mental causation per se); Nita Farahany, *A Neurological Foundation for Freedom*, STAN. TECH. L. REV. 4 (2012); Hedda Hassel Mørch, *The Evolutionary Argument for Phenomenal Powers*, 31 PHIL. PERSPECTIVES 293 (2018); Michael S. Pardo & Dennis Patterson, *Morse Mind and Mental Causation*, 11 CRIM L. & PHIL. 111–26 (2017); Jones & Wagner, *supra* note 8 (mildly neuroskeptical); Christopher P. Taggart, *Retributivism,* 111 (2017) (presenting "an alternative, rational-teleological account of what it means to explain actions in terms of mental states (or reasons)"); Katrina L. Sifferd, *What Does It Mean to Be a Mechanism? Stephen Morse, Non-reductivism, and Mental Causation*, 11 CRIM. L. & PHIL. 143 (2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512325; Christopher P. Taggart, *Retributivism, Agency, and the Voluntary Act Requirement*, 36 PACE L. REV. 645, 647 (2016) (making a case that "agent causation" is at least "plausible").

10. *See* Morse, *Inevitable Mind*, *supra* note 3, at 46–47 (recognizing the inconsistency and arguing that retribution is not "inherently harsh" and the alternatives are "disrespectful and dehumanizing").

I. MENTAL CAUSATION AND CRIMINAL JUSTICE

Modern criminal law takes the existence of mental causation for granted in a number of ways, presupposing that a person's conduct can be caused by her intentions, reasons and other mental states.[11] For example, the law's so-called "voluntary act" requirement—which requires a harm caused by volitional conduct for there to be a crime[12]—not only assumes that mental causation is real but accords it a central role in justifying punishment.[13] Similarly, the development and importance of mens rea requirements over the past 500 years[14] would have made no sense if it

11. The term "mental state" is used here, as in law, with its "ordinary language, common sense" meaning. Morse, *Folk Psychology*, *supra* note 3, at 2–3, 10; *see also* Morse, *Inevitable Mind*, *supra* note 3, at 35 ("Mental state requirements, including the mental states that are the criteria for voluntary action, mens rea, and justifications and excuses, reflect the criminal law's concern with intentionality and express the meaning of an action including the agent's attitudes towards the rights and interests of the victim"); Morse, *Translation*, *supra* note 3, at 530.

12. *See. e.g.*, Martin v. State, 17 So. 2d 427 (Ala. App. 1944); State v. Utter, 479 P.2d 946 (Wash. App. 1971); Rollin M. Perkins, *Rationale of Mens Rea*, 52 HARV. L. REV. 905, 912–13 (1939) (the harmful result must have been "brought about" by a voluntary act); WILLIAM BLACKSTONE, 4 COMMENTARIES ON THE LAWS OF ENGLAND § 2 (1758) ("[A]n unwarrantable act without a vicious will is no crime at all"); MODEL PENAL CODE §2.01(1). *See generally* Deborah W. Denno, *Crime and Consciousness: Science and Voluntary Acts*, 87 MINN. L. REV. 269, 275 (2002) ("Doctrinally, all criminal liability depends on one 'fundamental predicate': A defendant's guilt must be based on conduct and that conduct must include a 'voluntary act' or omission to engage in a voluntary act that the defendant was capable of performing"); DRESSLER, *supra* note 2, at 87 ("[A] person is not guilty of a crime unless her conduct includes a voluntary act"). Crimes by omission (failure to perform a legal duty to act, *see generally id.* at 105–11) also appear to be generally subject to a volition requirement, s*ee* United States v. Montague, 75 F. Supp. 2d 670 (S.D. Texas 1999), particularly inasmuch as omissions can be crimes only if the accused failed to act despite being "physically capable of performing the act." DRESSLER, *supra* note 2, at 105. The "volition" consists in voluntarily doing something other than performing the duty to act in question.

13. *See generally* Kevin W. Saunders, *Voluntary Acts and the Criminal Law: Justifying Culpability Based on the Existence of Volition,* 49 U. PITT. L. REV. 443 (1988) (discussing the criminal law's "reluctance to ascribe responsibility to a person whose body alone was involved in the act"). According to Holmes, an act is a "muscular contraction" that is "willed," and "[t]he reason for requiring an [willed] act is, that an act implies a choice, and that it is felt to be impolitic and unjust to make a man answerable for harm, unless he might have chosen otherwise." OLIVER WENDELL HOLMES, THE COMMON LAW 54 (Bos., Little Brown & Co. 1881).

14. *See, e.g.*, Elonis v. United States, 135 S.Ct. 2001, 2009 (2015) (stating that "the basic principle [is] that wrongdoing must be conscious to be criminal" and that "the 'general rule' is that a guilty mind is 'a necessary element in the indictment and proof of every crime'") (quoting United States v. Balint, 258 U.S. 250, 251 (1922)); Staples v. United States, 511 U.S. 600, 606 (1994) ("[O]ffenses that require no *mens rea* generally are disfavored"); United States v. U.S. Gypsum Co., 438 U.S. 422, 438 (1978); Morissette v. United States, 342 U.S. 246 (1952); *see also* United States v. Cordoba-Hincapie, 825 F. Supp. 485, 491 (E.D.N.Y. 1993) (stating that "the requirement of a guilty state of mind (at least for the more serious crimes) had been developed by the time of Coke" [1552-1634]) (citation omitted). *See generally* Francis Bowles Sayre, *Mens Rea*, 45 HARV. L. REV. 974, 975–1004 (1932) (magisterial treatment of the history of mens rea).

So-called "strict liability" crimes do not require mens rea but they are still subject to the requirement of a voluntary act (and hence seem to presuppose mental causation). *See* MODEL PENAL CODE §2.01(1); SUSAN MANDBERG, STRICT LIABILITY, THE ENCYCLOPEDIA OF CRIMINOLOGY AND

were not supposed that culpable mental states can play an active role as but-for causes[15] of unlawful conduct.[16]

CRIMINAL JUSTICE (2014) ("To convict a person of a strict liability crime, the prosecution still must prove a voluntary act or omission"); *see also* MICHAEL MOORE, PLACING BLAME: A THEORY OF THE CRIMINAL LAW 317 (2010) ("The voluntary act requirement requires that the accused intends that his body moves at all; the *mens rea* requirements are, respectively, that the accused . . . intends his movements to cause [the prohibited result, and knows they will cause the prohibited result].").

15. It should be noted that the discussion in this article concerns *causal* explanations. Not all explanations are causal explanations. For example, one might explain that birds build nests in the spring because they need places to raise their young (rather than in terms of the behavioral effects of season-induced hormones, etc.). This explanation is informative and adds to our coherent picture of the world because it portrays nest building as an instance of a larger phenomenon that is already understood, namely, of parents working to provide for offspring. But it is not *causal* explanation. Birds building nests may not have a clue that they are even about to have young, let alone what they will need. Similarly, even if mental states are epiphenomenal, particular kinds of mental states (for example, intentions to do X) may have a kind of explanatory utility as "mental ways of grouping physical states and events." W.V.O. QUINE, PURSUIT OF TRUTH 72 (1992). Since we do not have direct knowledge of our conative brain states as such, a person says "I did X because I intended to do X" rather than "I did X because of brain states that I experienced as intentions to do X" It would be analogous to describing the contents of my computer's memory by describing the words and figures that appear on its screen (as in: "I have pictures of my vacation stored on my laptop"). Notice, however, that while this analogy may be read to imply that the brain can detect and "read" the qualitative contents of mental states, it should not be so understood, for the reasons set out *infra* Part IV.B.

Following the philosopher of the mind Jaegwon Kim, the conception of "causation" is used in the present discussion to mean "actual productive/generative mechanisms involving energy flow, momentum transfer, and the like, and not merely . . . counterfactual dependencies." JAEGWON KIM, PHYSICALISM, OR SOMETHING NEAR ENOUGH 47 (2005) [hereinafter KIM, PHYSICALISM]. For example, even though on sunny days the hands on my watch move around the dial if and only if the sun crosses the sky (a counterfactual dependency), it would not be correct to say that my watch "causes" the sun to move across the sky, or vice versa. The "productive/generative" meaning of causation is, I believe, also the one that is understood in legal conception of "but-for causation."

16. The suggestion has been made that, strictly speaking, responsibility does not *require* a "but for" causal connection between culpable mental states and wrongful conduct, and that punishment can be justified as long as the culpable mental state was at least present during prohibited conduct; it is enough that wrongdoer mentally *wanted* to do the prohibited conduct or *approved* of doing it. *See* Nita A. Farahany, *A Neurological Foundation for Freedom*, *in* PHILOSOPHICAL FOUNDATIONS OF LAW AND NEUROSCIENCE 51, 66 (Patterson et al. eds., 2016) ("One can be morally responsible by choosing to act according to one's own desire to act, even if no other outcome would be possible"), adapting an argument originally made by Harry Frankfurt, *Alternative Possibilities and Moral Responsibility*, 66 J. PHIL. 829, 833, 837–39 (1969); Saunders, *supra* note 13, at 454 n.7, 466–75. A similar argument appears to have been made by Hilary Bok. *See* HILARY BOK, FREEDOM AND RESPONSIBILITY 203 (1995) ("[T]he claim that we are free is a claim not about the relation of our choices and actions to the theoretical system of causes but about their relation to the agent's practical system of reasons").

The idea that mens rea need only *coincide with* the criminal act and need not causally contribute to *producing* the crime does not, however, appear to have support in the law. It is, for example, at odds with the rule of accomplice liability under which a person cannot be held criminally accountable as an accomplice based solely on a secret wish to see the crime occur if the person does nothing to aid or encourage its actual commission. *See, e.g.*, State v. V.T., 5 P.3d 1234 (Utah Ct. App. 2000); State v. Hoselton, 371 S.E.2d 366 (W.Va. 1988). Kevin Saunders has made a worthy legal argument for punishing non-causative culpable mental states by analogy to the prohibition on criminal solicitation, which, under the Model Penal Code, is punishable even if the solicitation is not actually effectively communicated to the solicited. Saunders, *supra* note 13, at 466–75 (relying on MODEL PENAL CODE § 5.02). Arguably, however, this provision is inapposite if mental states are "inherently unlikely" to cause a crime, Saunders, *supra* note 13, at 474, because it presupposes that culpable mental states at

> The contention that an injury can amount to a crime only when inflicted by intention is no provincial or transient notion. It is . . . universal and persistent in mature systems of law . . . . A relation between some mental element and punishment for a harmful act is almost . . . instinctive . . . ."[17]

But even though the facticity of mental causation may seem intuitively obvious, there is reason for genuine doubt. The mental-causation hypothesis is founded on the belief that mental states, such as intentions and reasons, can cause changes to occur in the physical domain—and thereby affect and direct what people do. The theory is that mental states "play a genuinely causal role in explaining human behavior"[18] and, as such, are able to produce actions instead of, or at least in addition to, the biomechanical brain causation described by neuroscience.[19] The problem is that no one knows how.[20] It turns out that most or all of the obviousness of mental causation rests on a logical fallacy,[21] and "[c]enturies of philosophical effort have failed to establish that a mental act or volition may cause . . . a bodily movement."[22]

Punishment—the systematic infliction of hardship, deprivation and misery on human beings—is obviously a serious matter. As a criminal justice practice it is far out of step with the usual moral understandings of right and wrong.[23] For those who genuinely care whether such inflictions are morally right, a lot rides on the question of whether mental causation really occurs, viz. whether intentions and reasons really can cause physical conduct and influence behavior: "If the concept of mental causation that underlies folk psychology and current conceptions of responsibility is

---

least *can be* causally efficacious and thereby serve as a criterion of dangerousness. *See* DRESSLER, *supra* note 2, at §28.01[D] and comments to the Model Penal Code quoted in Saunders, *supra* note 13, at 469 n.109, 472. It seems to me, however, that all of these arguments are subject to the same basic objection: If culpable mental states are causally inefficacious, then they are harmless, and it is illogical to punish a person for an aspect of her personality that is harmless if she would not be subject to punishment for another aspect of her personality that *is* harmful but not in itself culpable.

17. *Morissette*, 342 U.S. at 250–51.

18. Morse, *Translation*, *supra* note 3, at 532.

19. Morse, *Folk Psychology*, *supra* note 3, at 1; Morse, *Translation*, *supra* note 3, at 530.

20. Morse, *Inevitable Mind*, *supra* note 3, at 33 ("[W]e do not have a clue").

21. *See infra* Part IV. Specifically, the main case for mental causation rests on the fallacy of "post hoc ergo proper hoc," the false assumption that a given event was caused by a prior event just because it came after the prior event.

22. Saunders, *supra* note 13, at 466.

23. Just to be clear, nothing in the neuroscience explanation of human behavior denies the important moral distinction between right and wrong conduct ("normativity") or implies that such a distinction cannot be an effective guide for behavior. For my earlier brief exploration of these issues, *see* John A. Humbach, *Does Hard Incompatibilism Really Abolish 'Right' and 'Wrong'? Some Thoughts in Response to Larry Alexander*, (Mar. 14, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2933749; *see also infra* Part II.D.

false, our responsibility practices . . . would appear unjustifiable."[24] Modern punishment precepts and methodologies—if not the whole project of criminal "justice"—would be thrown into question. And it turns out, on reflection, that the case for mental causation could hardly be more flimsy.

Advances in neuroscience since the middle of the last century have greatly expanded our understanding of the biomechanical processes occurring in the brain and central nervous system.[25] They have revealed in considerable detail how electrical activity in the brain can cause bodily movements and therefore produce behavior, not just in human beings but in virtually every kind of multicellular animal organism found on earth.[26] They show how biological organisms function as complex adaptive input-output systems in which electrochemically encoded information acquired by the sense organs and internal sensors is passed to the brain where various populations of neurons computationally process the information and integrate it with neuronally-embodied information stored from past experiences, on the basis of which they then produce the goal-directed, coordinated motor impulses that actuate the muscles in response to surrounding circumstances.[27] This is all it takes to cause every move we make.

In other words, the evidence of neuroscience shows that the causes of human behavior traditionally thought to be "mental" (including behavior punished by the criminal law) are almost certainly entirely physical in nature, occurring in strict accordance with the physical laws that govern the electrochemistry within nerve cells, the brain and the body.[28] Unlike

---

24. Morse, *Translation*, *supra* note 3, at 543*; see also* Morse, *Inevitable Mind*, *supra* note 3, at 46–47; Morse, *Folk Psychology*, *supra* note 3, at 2–3.

25. *See supra* notes 4 and 6.

26. In a comment on the fundamental similarities of the sources of brain-directed behavior, Owen Jones and Timothy Goldsmith have written: "In all but a few universities, human behavior is studied . . . in one set of buildings, while the behavior of every [other] species . . . is studied in other buildings. There are reasons for this—but few good ones." Owen D Jones & Timothy H. Goldsmith, *Law and Behavioral Biology*, 105 COLUM. L. REV. 405, 407 (2005).

27. *See* SAPOLSKY, *supra*, note 6, at 21–77, 535–36 (describing and summarizing how the brain chooses and produces bodily movements); PASSINGHAM, *supra*, note 4, at 66–81; and other sources cited *supra* note 6.

Note that saying the brain produces behavior by biomechanical or "computational" processes is not at all the same as saying the workings of the *mind* are "computational" or embracing the so-called computational theory of mind. It is important that the two not be confused. In saying the brain produces behavior by biomechanical or computational processes, I mean only that the workings of individual neurons and the connections between them follow simple algorithms and that the combined outputs of the algorithmic neurons in their networks produce the motor outputs of the brain.

28. Notice that to be in strict accordance with physical laws is not necessarily the same as to be deterministic. Physical interactions are themselves sometimes indeterministic (notably at the quantum level) and there is at least some reason to believe that neuronal activity may not always be deterministic. *See* Fileva & Tresan, *supra* note 8, at 10; Adina Roskies, *Neuroscientific Challenges to Free Will and Responsibility*, 10 TRENDS IN COGNITIVE SCI. 420 (2006) ("Whether or not a neuron will fire, what pattern of action potentials it generates, or how many synaptic vesicles are released

the "folk psychology" of mental causation that underlies modern criminal law,[29] the neuroscience explanation of behavior is not merely consistent with the physical laws that apply throughout the Universe. It is built upon them. Despite extensive studies of the complex of mechanisms within the person that produce human behavior,[30] no empirical evidence has ever appeared that alternative, non-neuronal sources of coordinated muscular activation (behavior) might exist. No explanatory gaps have been encountered in the neuroscience explanation of behavior that could be filled by mental causation.

This is not to say there are no remaining gaps in the scientific understanding of human psychology. There are, and neuroskeptics point to these epistemic lacunae as reasons to discount the relevance of neuroscience to criminal justice.[31] One alleged "major problem" bearing on that relevance, identified by neuroscientist Robert Sapolsky, is that the science to date is very long on explanation but short on ability to specifically *predict* misconduct, mostly due to the "multifactorial" nature of behavioral biology.[32] Every bodily movement is, he points out, produced by an interplay of huge numbers of causal factors,[33] and this multiplicity tends to make the predictions of neuroscience only approximate and "statistical."[34] Professor Sapolsky may however be unduly modest: An inability to get a precise reading or measurement of a multiplicity of causal factors is not at all the same as not knowing what

---

have all been characterized as stochastic phenomena in our current best models"). In other words, it is possible that "neurodeterminism," in the sense of control of the body by neuronal activity, is not fully "deterministic" (in the super-hard sense that every state of the Universe is ineluctably entailed by preceding states). But even if neurodeterminism is not fully deterministic, this would not in itself mean that the operations of the brain and central nervous system leave room for a causal role to be played by "free will," mental causation or other alternatives to neuronal activity in controlling what people do.

29. *See generally* Morse, *Folk Psychology*, *supra* note 3, at 1; Morse, *Translation*, *supra* note 3, at 530; Morse, *Inevitable Mind*, *supra* note 3, at 34.

30. *See supra* note 4.

31. *See, e.g.*, Stephen J. Morse, *Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty*, 62 MERCER L. REV. 837, 859 (2011); Dennis Patterson, *Neuromania: A Review of Peter A. Alces, The Moral Conflict of Law and Neuroscience*, 5 J. L. & BIOSCI. 440, 441 (2018) ("[N]euroscience simply has not progressed to the point where it can even tell us how the brain enables the mind"); Uri Maoz & Gideon Yaffe, *What Does Recent Neuroscience Tell Us About Criminal Responsibility?* 3 J. LAW BIOSCI. 120 (2015).

32. SAPOLSKY, *supra* note 6, at 598–605.

33. *See id.* Professor Sapolsky discusses the many kinds of factors influencing behavior during the second before it occurs, minutes before, hours before, etc. Just to begin to give an idea of the numerosity of factors, they potentially include the specific synaptic strengths and connections of millions of neurons that may figure into the balance leading to a given behavioral choice—not to mention interoceptive and hormonal inputs from outside the brain that potentially implicate virtually any part of the body's physiology. And these are just factors existing during the seconds before the movement occurs. *Id.* at 21–98.

34. *See id.* at 36, note †.

those factors are—and it is surely not a good reason to conclude that something other than ordinary physical forces must therefore be at work. It is similarly not possible to get a reading on all of the complex factors that make the weather or move the stock market, but no one thinks this is a reason to believe that immaterial spirit-like forces are involved. That sort of logical leap is seen only among friends of mental causation.

Perhaps more notably, neuroscience still does not provide a solution to the "really hard problem of consciousness,"[35] which is to explain how mental states and consciousness could possibly exist in a physical universe in the first place.[36] "Despite the astonishing advances in neuroimaging and other neuroscientific methods, we still do not have sophisticated causal knowledge of how the brain enables the mind and action generally."[37] But this argument against the relevance of neuroscience also misses the mark. When the question is how well neuroscience explains *behavior*, it is ultimately beside the point whether it is able to explain something *else,* namely consciousness, until there is reliable evidence that consciousness has something to do with behavior. What the findings of neuroscience do provide is a sophisticated knowledge of the mechanisms by which the brain produces bodily movements, and it is *that* knowledge which is relevant to a causal explanation of behavior: The evidence shows that unbroken sequences of computational neuronal events run from sensory inputs to motor outputs, and these neuronal events are the sole determinants *in evidence* of what human beings and other innervated organisms do. No one has ever seen the slightest physiological evidence that there is an alternative or parallel "mental" path of behavior causation. And until research uncovers evidence that mental states or consciousness plays a role in producing human behavior, an explanation of how mental states come into being, though intensely interesting as a philosophical matter, is simply not necessary for a complete causal explanation of why people do what they do.

---

35. David J. Chalmers, *Facing Up to the Problem of Consciousness*, 2 J. CONSCIOUSNESS STUD. 200 (1995).

36. *See* Robert Van Gulick, *Consciousness* 5.2, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2014); *see also* THOMAS NAGEL, MIND AND COSMOS: WHY THE MATERIALIST NEO-DARWINIAN CONCEPTION OF NATURE IS ALMOST CERTAINLY WRONG (2012); DAVID J. CHALMERS, THE CONSCIOUS MIND: IN SEARCH OF A FUNDAMENTAL THEORY (rev. ed. 1997); Hedda Hassel Mørch, *Is Matter Conscious?*, NAUTILUS (Apr. 6, 2017), http://nautil.us/issue/47/consciousness/is-matter-conscious.

37. Stephen J. Morse, *Law and the Sciences of the Brain/Mind*, *in* OXFORD HANDBOOK ON LAW AND THE REGULATION OF TECHNOLOGY 24 (Roger Brownsword, ed., 2017) [hereinafter Morse, LAWSCI], https://scholarship.law.upenn.edu/faculty_scholarship/1642/; Morse, *Inevitable Mind*, *supra* note 3, at 33 ("[W]e do not have a clue"); LEGAL, MORAL, AND METAPHYSICAL TRUTHS: THE PHILOSOPHY OF MICHAEL S. MOORE 233, 235 (Kimberly Ferzan & Stephen J. Morse eds., 2015) ("The brain enables the mind and action, but we have no idea how"); *see also infra* note 161.

The idea that the mind causes behavior has been long accepted by most people and by the law despite the lack of evidence for it.[38] But this acceptance is understandable when one considers that, until recently, no other causal explanation of human behavior has had any evidentiary support either. As long as the mental-causation model had no non-speculative challenger, nothing stood in the way of its becoming firmly embedded in popular notions of justice as well as in the law itself; it was able to persist unquestioned essentially by default. Recent neuroscience research has, however, changed this picture significantly. The emergence of a documented, evidence-based alternative to the mental-causation hypothesis has deprived the folk-psychology narrative of its explanatory monopoly. By providing a physical account of behavior causation, from sensory input to muscular output, that leaves no gaps that mental states are needed to fill,[39] the neuroscience alternative to folk-psychology has begun to make mental causation look, at best, like an otiose excrescence on an otherwise elegant explanatory picture.

The emergence of a creditable alternative to mental causation as a causal explanation of behavior inevitably suggests a need for a reevaluation and corresponding rethinking of criminal justice practices and punishment criteria that presuppose that mental causation occurs.[40] If indeed blame and responsibility are based on mental causation, the question arises whether it is still reasonable to think[41] that those who do wrong "deserve" to receive suffering and deprivation purposely inflicted by the state.

Given the formidable neuroscience findings as to the causes of human behavior, why do so many still look with a skeptical eye at the findings' implications for law? The most probable reason is that neuroscience's

38. *See infra* Part IV.

39. *See supra* note 27.

40. *See, e.g.*, Greene & Cohen, *supra* note 8; *cf.* Morse, *Translation*, *supra* note 3, at 534 ("The law will be fundamentally challenged only if neuroscience or any other science can conclusively demonstrate that the law's psychology is wrong and we are not the type of creatures for whom mental states are causally effective").

41. By the way, I do not mean to express any particular philosophical commitments by making casual use of everyday mentalistic words like "think." Because I am writing in English, there is little choice but to use such locutions and expressions since the English language was, for better or worse, devised by people who believed in mental causation and it is, accordingly shot through with marks of that belief. Such word choices should not, however, be taken as an implicit acquiescence in any particular philosophy of the mind, such as mental states are anything but epiphenomenal (i.e., causally inert). Accordingly, consistent with the neuroscience explanation of behavior, these casual uses of mentalist locutions should (unless the context otherwise requires) be understood to refer to the brain states and neuronal activity on which our mental life and consciousness depend. *See* Morse, *Translation*, *supra* note 3, at 532 ("[M]ental states . . . are fully produced by and realizable in the brain").

biomechanical explanation of behavior challenges a deeply felt belief—the belief that lawbreakers *deserve* to suffer because their intentions cause harmful acts.[42] It contradicts the widely-shared urge to deal with harmful events by finding someone to blame and meting out harm in return, instead of seeking out causes and implementing prevention. Those who doubt the justice implications of neuroscience do not, however, usually contest the validity of research findings themselves. As a practical matter, only a neuroscientist is well equipped to do that, and there are few if any actual neuroscientists who do not broadly share the scientific biomechanical view of human behavior. For non-scientists, the expedient thing to do is in any case to argue, not that the findings of neuroscience are wrong, but rather that they are simply irrelevant: They are irrelevant to normative questions generally[43] and to criminal justice policy in particular because they do not rule out mental causation.[44] It is this skeptical stance that will be the primary focus of the discussion that follows.

Professor Stephen J. Morse is a leading exponent of the view that nothing in the findings of neuroscience requires us to rethink the prevailing conceptions of moral responsibility and criminal justice practice.[45] His writings,[46] taken together, present what is probably the most comprehensive and thoughtful statement by a legal scholar of the position

---

42. *See* Morse, *Neurolaw*, *infra* note 46, at 16 ("This thought [that responsibility and, hence, 'just deserts' may not be justified] is what disturbs people about scientific understanding of human behavior, which relentlessly exposes the numerous causal variables that seem to toss us about like light ships in a raging sea storm").

43. *See infra* Part II.D.

44. *See infra* Part II.C.

45. *See* SAPOLSKY, *supra* note 6, at 598 ("He is the definitive advocate of free will being compatible with a deterministic world").

46. While the list of Professor Morse's publications in this area is uncommonly long, many of them cover much the same ground and the present discussion will draw primarily from those cited *supra* note 3 and the following: Stephen J. Morse, *NeuroEthics: NeuroLaw* OXFORD HANDBOOK ONLINE (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2919011 [hereinafter Morse, *Neurolaw*]; Morse, LAWSCI, *supra* note 37; Morse, *Translation*, *supra* note 3, at 530; Morse, *Inevitable Mind*, *supra* note 3; Stephen J. Morse, *Neuroscience, Free Will, and Criminal Responsibility*, *in* FREE WILL AND THE BRAIN: NEUROSCIENTIFIC, PHILOSOPHICAL, AND LEGAL PERSPECTIVES 283 (Walter Glannon ed., 2015) [hereinafter Morse, *Free Will*]; Stephen J. Morse, *Criminal Law and Common Sense: An Essay on the Perils and Promise of Neuroscience*, 99 MARQUETTE L. REV. 39 (2015) [hereinafter Morse, *Common Sense*]; Stephen J. Morse, *Neuroscience and the Future of Personhood and Responsibility* (2013), https://www.brookings.edu/wp-content/uploads/2016/06/0203_neuroscience_morse.pdf; Stephen J. Morse, *Moore on the Mind* 4 (2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2705192 [hereinafter Morse, *Moore on the Mind*]; Stephen J. Morse, *Scientific Challenges to Criminal Responsibility*, *in* JOEL FEINBERG ET AL., PHILOSOPHY OF LAW (9th ed. 2014) [hereinafter Morse, *Scientific Challenges*]; Stephen J. Morse, *Brain Overclaim Redux*, J. LAW & INEQUALITY (2012) [hereinafter Morse, *Overclaim Redux*]; Stephen J. Morse, *Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty*, 62 MERCER L. REV. 828 (2011) [hereinafter Morse, *Exuberance*]; Stephen J. Morse, *Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic* 3 OHIO ST. J. CRIM. L. 397–412 (2006) [hereinafter Morse, *Overclaim*].

that "neuroscience has little to contribute to a more just and accurate criminal law decision-making concerning policy, doctrine, and individual case adjudication."[47] This paper undertakes to critique[48] the skeptical view of neuroscience's value to the quest for justice, with particular attention to arguments made by Professor Morse as the leader in the field. Most of the points made in the discussion should, however, apply with similar force to any theories or doctrines of personal responsibility that are based on the supposition that human actions are caused by intentions or influenced by reasons.[49] My conclusion is that the findings of neuroscience, though incomplete, already present an extremely strong likelihood that the currently dominant moral justifications for punishment inflictions are based on erroneous factual assumptions.

II. THE CORE OF THE SKEPTICAL VIEW

The central thesis of the neuroskeptical position is that the relevance of neuroscience to criminal law and punishment practices is vastly exaggerated and "overclaimed."[50] "Given how little we know about the brain-mind and brain-mind-action connections," Professor Morse has written, "to claim that we should radically change our conceptions of ourselves and our legal doctrines and practices based on neuroscience is a form of neuroarrogance."[51]

---

47. Morse, *Exuberance*, *supra* note 46, at 859.

48. I refer the reader also to the thoughtful critique of Professor Morse (and his neuroenthusiast counterparts, Joshua Greene and Jonathon Cohen, *supra* note 8). *See* Adam Kolber, *Will There Be a NeuroLaw Revolution?* 89 IND. L. REV. 807, 811 (2014). Michael Pardo and Dennis Patterson have also critiqued Professor Morse's reliance on the mental-causation hypothesis, suggesting instead that responsibility and punishment can be based instead on the teleological *explanation* for the conduct in question. Michael S. Pardo & Dennis Patterson, *Morse, Mind and Mental Causation*, 11 CRIM. L. & PHIL. 111 (2017). The problem with this alternative is that a teleological explanation is not a causal explanation unless the teleological "reason" for the behavior is represented in the agent, mentally or physically, prior to the action that it is supposed explain/cause. This leads us back again to the same question faced by the standard mental-causation hypothesis. If the teleological "reason" is represented solely as a physical state, how can it have moral significance; if it is represented as a mental state, how can it have causal efficacy? This is important because, if the culpable teleological explanation does not play a causal role in producing the behavior, then we would again have a case of blame being based on the mere *presence* of a culpable mental state that is, apart from its simultaneity, unconnected to the behavior in question. *See supra* note 16.

49. *See supra* note 9.

50. *See* Morse, *Overclaim*, *supra* note 46, at 397–412; *see also* Morse, *Neurolaw*, *supra* note 46, at 7, 31; Morse, *Exuberance, supra* note 46, at 859.

51. *Id.*

A. Mental Cause, Agency and Responsibility

The "conception of ourselves" that neuroskepticism seems most concerned to preserve is the idea that people are morally responsible "agents," which means "creatures who act intentionally for reasons, who can be guided by reasons."[52] The assumption that persons are "agents," and are therefore responsible for their acts, is thought to underlie the moral rightness of key legal doctrines and practices such as punishment and the notion that offenders "deserve" it.[53] "If practical reason plays no role in explaining our behavior, as some neuroscientists . . . claim, current responsibility doctrines and practices would have to be radically altered or jettisoned altogether."[54] Mental causation by intentions and reasons, which is what defines a person as an agent, is thus the key to moral responsibility and just deserts. "It is an incoherent notion to have genuine responsibility without agency."[55]

By showing that "behavior can be explained solely by reference to physical states . . . and the laws governing changes in those physical states,"[56] neuroscience presents a physiological alternative to the mental-causation explanation of behavior. This alternative explanation threatens a fundamental premise of current criminal-justice doctrines and practices. Although Professor Morse concedes that "the brain, the final pathway to action, is nothing but a mechanism"[57] and "[m]achines do not deserve . . . punishment,"[58] he insists that human beings are not just machines but agents. And as agents, offenders are morally responsible and deserve punishment precisely *because* they are agents: The brain may be a "mechanism" but, he maintains, "[h]uman behavior cannot be adequately understood if mental state causation is completely excluded or eliminated."[59] "Behavior," he says, "is causally explained by *mental states*

52. Morse, *Inevitable Mind*, *supra* note 3, at 33; Morse, *Common Sense*, *supra* note 46, at 40. Notice that the concept of "agent" and "agency" in this context is not the same as its usual meaning among lawyers. Rather, the word agent is used here in a philosophical sense to refer to an individual who is capable of acting based on intentions and for reasons.

53. Morse, *Inevitable Mind*, *supra* note 3, at 32–34; Morse, *Folk Psychology*, *supra* note 3, at 19 ("[I]f humans are not conscious and intentional creatures who act for reasons that play a causal role in our behavior, then the foundational facts for responsibility ascriptions are mistaken").

54. Morse, *Folk Psychology*, *supra* note 3, at 2–3 (If "[n]euroscientific discoveries . . . demonstrate that mental states do not causally explain our behavior . . . it provides another, independent ground for the claim that responsibility is impossible").

55. Morse, *Common Sense*, *supra* note 46, at 69; Morse, *Inevitable Mind*, *supra* note 3, at 46.

56. Andrew Eshleman, *Moral Responsibility*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2014), https://plato.stanford.edu/entries/moral-responsibility/.

57. Morse, *Neurolaw*, *supra* note 46, at 16 ("Neuroscience . . . exposes that the brain, the final pathway to action, is nothing but a mechanism").

58. Morse, *Inevitable Mind*, *supra* note 3, at 34.

59. Morse, *Translation*, *supra* note 3, at 532.

such as desires, beliefs, plans, willings, and intentions."[60] In other words, mechanical brain or not, as long as people have mental states that are involved in causing their behavior, offenders can be held morally responsible as agents and deserve to be punished for their wrongs.

Somewhat surprisingly, perhaps, the neuroskeptical insistence that mental states can cause behavior does not necessarily mean a rejection of *determinism* —roughly, "the idea that every event is necessitated by antecedent events and conditions together with the laws of nature."[61] Professor Morse calls determinism a "plausible working hypothesis,"[62] and one can see why he would want to do so. Today, after all, determinism is "a view accepted by most scientifically-informed people even if they are also humanists,"[63] and to insist that events can happen other than according to natural laws is to risk putting oneself outside the bounds of contemporary mainstream scholarship. But this reluctance to reject determinism seems to land Professor Morse in a bit a contradiction: He tells us on one hand that "[b]ehavior is causally explained by mental states" such as desires, beliefs intentions, etc.[64] But then he says, deterministically, that "we inhabit a thoroughly material, physical universe in which all phenomena [including, presumably, human behavior] are caused by physical laws,"[65] apparently conceding that people's conduct is not necessarily "free," i.e., "uncaused by anything other than themselves."[66] The latter statements, though routinely consistent with determinism, make it sound a lot like people are *not* agents. The challenge that is faced by neuroskeptics like Professor Morse is to find a "plausible,

60. Morse, *Inevitable Mind*, *supra* note 3, at 34 (emphasis added); *see also* Morse, *Translation*, *supra* note 3, at 532. Elsewhere, Professor Morse has said that folk psychology "insists only that human action is *in part* causally explained by mental states," Morse, *Translation*, *supra* note 3, at 530 (emphasis added), but makes clear that it is "virtually" only the mentally caused part for which "agents deserve to be praised, blamed, rewarded, or punished." Morse, *Translation*, *supra* note 3, at 532; Morse, *Exuberance*, *supra* note 46, at 841.

61. Carl Hoefer, *Causal Determinism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2016), https://plato.stanford.edu/entries/determinism-causal/#ChaDet. Professor Morse describes determinism as, roughly, the idea "that all events have causes that operate according to the physical laws of the universe and that were themselves caused by those same laws operating on prior states of the universe in a continuous thread of causation going back to the first state"—with the usual (essentially irrelevant) allowance for quantum-level events. Morse, *Neurolaw*, *supra* note 46, at 17.

62. Morse, *Inevitable Mind*, *supra* note 3, at 45.

63. Morse, *Neurolaw*, *supra* note 46, at 16 & 17.

64. Morse, *Inevitable Mind*, *supra* note 3, at 34 (emphasis added); *see also* Morse, *Translation*, *supra* note 3, at 532. As discussed *infra* Part IV.A., the contradiction may turn out on closer analysis to be a bit different from what it first seems, lying more in Professor Morse's adamant rejection of neurodeterminism and not in determinism *tout court*.

65. *See* Morse, *Common Sense*, *supra* note 46, at 47; Morse, *Inevitable Mind*, *supra* note 3, at 33.

66. I.e., that determinism may be true. Morse, *Neurolaw*, *supra* note 46, at 16. *Id.* at 17–18 (stating that no such "panicky" metaphysics is necessary).

mainstream" way to embrace determinism, at least provisionally, without being stuck with its logical implication that persons are not agents and that "genuine responsibility" is thus impossible.[67]

Like many if not most modern theorists,[68] Professor Morse tries to thread this needle by declaring adherence to a long established, have-your-cake-and-eat-it-too position known as *compatibilism*,[69] "a set of similar theories that hold with varying intensity that responsibility is possible in a deterministic universe as long as *agents* have the capacity to act according to their reasons"[70] and intentions.[71] This compatibilism, which Professor Morse calls "the dominant position among philosophers of responsibility,"[72] allows one to accept determinism as a "plausible working hypothesis"[73] (and to demur on the fraught question of whether wrongdoers "can do otherwise"[74]) even while firmly denying that physical brains alone determine what a person does in a given situation: "We are not Pinocchios," he writes, "and our brains are not Geppettos pulling the strings."[75]

---

67. Morse, *Neurolaw*, *supra* note 46, at 13 ("It is an incoherent notion to have genuine responsibility without agency"); *see also* Morse, *Common Sense*, *supra* note 46, at 67; Morse, *Inevitable Mind*, *supra* note 3, at 46.

68. Morse, *Inevitable Mind*, *supra* note 3, at 45; Morse, *Overclaim*, *supra* note 46, at 402.

69. For a summary of many of the varieties of compatibilism, *see* Michael McKenna, *Compatibilism: State of the Art*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2015), http://plato.stanford.edu/entries/compatibilism/supplement.html.

70. Morse, *Inevitable Mind*, *supra* note 3, at 45; *see also* Morse, *Neurolaw*, *supra* note 46, at 17 ("For those who adopt some variant of this position, agents may be responsible if, roughly, they act intentionally, with reasonably integrated consciousness, suffer from no major rationality defects, and act free of compulsion.").

71. Morse, *Neurolaw*, *supra* note 46, at 18. It is a common strategy of modern compatibilism to posit various features or elements of the reasoning process as appropriate bases for ascribing moral responsibility. Michael McKenna, *Compatibilism: State of the Art*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2009), http://plato.stanford.edu/entries/compatibilism/supplement.html. For example, Hilary Bok (among others) posits our use of 'practical reasoning' as the place in which to find a basis for ascribing moral responsibility. HILARY BOK, FREEDOM AND RESPONSIBILITY 203 (1995) ("Those actions that reflect our will, and with respect to which the question what reason we had for performing them can arise, are actions that . . . we are morally responsible for performing . . . ."). For Daniel Dennett, it is using our evolved capacity to reflect and adapt. DANIEL C. DENNETT, FREEDOM EVOLVES 268 (2003) ("[O]ur reflections will actually help *determine* which trajectory our future holds"(!); emphasis added). For R. Jay Wallace, the basis of responsibility is the rational power "to grasp and apply moral reasons, and the power to control one's behavior in the light of such reasons." R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS 7 (1994).

72. Morse, *Inevitable Mind*, *supra* note 3, at 45; Morse, *Overclaim*, *supra* note 46, at 402.

73. Morse, *Inevitable Mind*, *supra* note 3, at 45.

74. Morse, *Common Sense*, *supra* note 46, at 49; *see also* Morse, *Inevitable Mind*, *supra* note 3, at 45 ("[C]ontra-causal freedom is simply not necessary"). *See* quotation *infra* note 78.

75. Morse, *Inevitable Mind*, *supra* note 3, at 49. Adam Kolber has written that a "fundamentally compatibilist criminal law would be insulated from advances in neuroscience that merely elucidate brain mechanism." Kolber, *NeuroLaw Revolution?*, *supra* note 48, at 822. I am not so sure, especially if (as seems to be happening) the belief in mental causation appears increasingly to be improbable as the "best explanation" of the growing body of data. *See* Part V.

The neuroskeptics' urgent moral concern about who or what is "pulling the strings" of criminal conduct is, to be sure, a relatively modern one. At one time, requital alone—repaying harm with harm in return—sufficed in itself to justify punishment.[76] If requital still sufficed today, none of the concern about who pulls the strings would be necessary. Punishment could simply be a per se consequence of doing harm, and that would be that. Today, however, it is generally considered morally obtuse to punish people for harms they utterly did not intend.[77] Most would probably say that a person who faints and knocks another off a dock does not deserve hardship and deprivation at the hands of the state for the harm she has "done." Indeed, it is likely the case that most people in legal circles now accept that "the criminal deserves punishment *because* he could have acted differently,"[78] but voluntarily chose not to.[79] It is said, for example, that harm caused by a person in an automatonistic state (moving as an automaton) does not deserve punishment.[80] More generally, most people today would probably say that punishment can be justly imposed only on persons who commit intentional or, at least, voluntary wrongdoing despite having the freedom to do otherwise.

---

76. *See* Perkins, *supra* note 12, at 906; FRIEDRICH NIETZSCHE, ON THE GENEALOGY OF MORALS 62–63 (Walter Kauffmann trans., 1967) (1887) (discussing the moral distinction between punishment for *harm* alone, akin to repayment of a debt, as opposed to punishment only if there was also *fault*; emphasis added).

77. *See* Morissette v. United States, 342 U.S. 246, 250 (1952) (citing "the child's familiar exculpatory 'But I didn't mean to'" as "instinctive"). Curiously, however, there is evidence that personal attitudes about punishment and its appropriateness may be less a "given" and more pliable than we may think, being contingent on specific neuronal activity that is subject to modification by mechanical means, such as magnets applied outside the skull (repetitive transcranial magnetic stimulation). *See* Jones & Wagner, *supra* note 8, at 7; Matthew Ginther et al., *Parsing the Behavioral and Brain Mechanisms of Third-Party Punishment*, 36 J. NEUROSCIENCE, 9420–34 (2016). Interesting experimental results by Shaun Nichols who found that, when people are persuaded that determinism is true, they are more likely to say that wrongdoers are morally responsible than when they merely *imagine* that determinism is true. Shaun Nichols, *Experimental Philosophy and the Problem of Free Will*, 331 SCIENCE 1401, 1402 (2011). In other words, when philosophers and others who believe in free will engage in thought experiments and speculations about the putative effects of determinism beliefs on attributions of responsibility, their conclusions may be fatally infected by their own personal beliefs that determinism is actually false and free will is true.

78. NIETZSCHE, *supra* note 76, at 62–63. *See* James W. Moore, *What Is the Sense of Agency and Why Does it Matter?* FRONTIERS IN PSYCHOLOGY 7 ("[F]or most people it only makes sense to hold someone responsible for their actions if they are freely in control of them."); HOLMES, *supra* note 13, at 54 ("[It is felt to be impolitic and unjust to make a man answerable for harm, unless he might have chosen otherwise"). See also Honoré's excellent analysis of "can/can't do otherwise" concept in A.M. Honoré, *Can and Can't*, 73 MIND 463 (1964), concluding that to say what one means by "can't do otherwise" leads into an infinite regress, which can be escaped only by making assumptions that usually cannot be justified. Daniel Dennett presents an analysis that is similar except it reaches the opposite conclusion concerning responsibility by seemingly ignoring the infinite regress. DENNETT, *supra* note 71.

79. *See supra* notes 12–13 and accompanying text.

80. *See* State v. Utter, 479 P.2d 946 (Wash. App. 1971); MODEL PENAL CODE §2.01(1).

Until recently, a concern to reconcile determinism with rejection of pure requital-based punishment might have tempted a neuroskeptical compatibilist to search for a modified definition of "free will." One might reason, for example, that even if determinism is true and persons cannot "do differently," they still have enough "free will" to be held responsible and deserve punishment as long as nothing prevents them from acting on the intentions and reasons they have: As long as one is free to do what one wants to do, it does not matter (for responsibility and punishment) that one is not free to choose what one wants.[81] But is this really "free will" or just freedom of movement? The point is arguable, but there has always been something fishy about the efforts of compatibilists to establish that free will exists by establishing instead that something else exists and then calling that something else "free will."[82]

Professor Morse has, however, found a way to elide the long-inconclusive "free will" debate.[83] Pointing out that "the law's criteria for responsibility . . . are essentially behavioral—acts and mental states,"[84] he argues that legal doctrine does not actually care whether defendants "could have acted differently" or not.[85] All the law requires, at least for most

81. Morse, *Inevitable Mind*, *supra* note 3, at 45; *see also supra* note 71.

82. Writing in 1788, Kant famously called the compatibilist's casuistry "a wretched subterfuge" and "word-jugglery." IMMANUEL KANT, THE CRITIQUE OF PRACTICAL REASON Ch III (Thomas Kingsmill Abbott trans., 2014) (1788)

83. Morse, *Neurolaw*, *supra* note 46, at 16; *Inevitable Mind*, *supra* note 3, at 45 ("People have been arguing for centuries about whether free will and responsibility are possible . . . . Neuroscience adds nothing to this debate"); *see also* Owen D. Jones, *the End of (Discussing) Free Will*, THE CHRONICLE REVIEW (Mar. 23, 2012), at 89, *reprinted in* OWEN D. JONES ET AL., LAW AND NEUROSCIENCE 132 (2014).

84. Morse, *Free Will*, *supra* note 46, at 259.

85. Morse, *Folk Psychology*, *supra* note 3, at 3–13 ("[F]ree will or the lack of it is not a criterion for criminal responsibility or non-responsibility"); Morse, *Common Sense*, *supra* note 46, at 53; Stephen J. Morse, *The Non-Problem of Free Will in Forensic Psychiatry and Psychology*, 25 BEH. SCI. L. 203 (2007) ("[F]ree will or its lack is not a criterion for any legal doctrine"). Although Professor Morse is obviously aware that the law generally requires, in order for conduct to be a crime, that is must include a "voluntary act," discussed *supra* notes 12–13, it is at least arguable that the voluntary-act requirement is not a requirement of free will. For example, an act can be "voluntary" (i.e., pursuant to a person's volition) even if the volition itself was utterly dominated by an outside force. Suppose, for example, that A strongly desires to throw a rock at B because an evil genius with a wireless mind-controller—or the deterministic operations of cause and effect—made A have that desire. If nothing prevented A from doing what he strongly desired ("willed") to do, his act could be considered "voluntary" even though his volition was ineluctably caused and molded by the evil genius. More prosaically, a person can be guilty of a crime that she "willfully" committed, even though she chose to commit it due to a credible threat of immediate death as long as the death threat did not happen to legally qualify as "duress." *See* People v. Anderson. 28 Cal. 4th 770 (2002) (holding that duress is not a defense to murder); United States v. Contento-Pachon, 723 F.2d 691 (9th Cir. 1984) (discussing the possibility of dire threats that may not be duress due to, for example, non-imminence). That said, however, I find it highly dubious that the common law judges over the centuries, imbued as they would have been with folk psychology, ever actually conceived of "voluntary acts" as being about anything *but* free will, specifically, the idea that indeterministic free will lies at the heart of human action. *See generally* Kolber, *NeuroLaw Revolution?*, *supra* note 48, at 823–26 ("soul-based

serious crimes, is that defendant did the prohibited act with a culpable mental state; there is no additional requirement that mental state be under her control or otherwise "free." There is, in other words, no *legal* requirement of free will[86] and, that being the case, the neuroskeptic can argue that legal responsibility can attach and punishment be deserved based on agency alone. All that is required for conviction is that the causes of a defendant's conduct include mental states, such as intentions, mixed into the chain of physical causes. In other words, by replacing free will with mental causation as the defining criterion of responsibility and just deserts, it is possible to bypass the whole free-will quagmire and simply concede that determinism is "plausible." This move also dissolves the apparent contradiction noted above. After all, there is nothing about mental causation that makes it necessarily inconsistent with determinism.[87] People's causative intentions and reasons can be just as much pre-determined by other causes as non-mental causal factors can be. So nothing prevents the neuroskeptical compatibilist from accepting the truth of determinism. The only kind of determinism that a neuroskeptical compatibilist can *not* accept is *neuro*determinism, at least not the version implied by neuroscience that denies a causal role for mental states on the ground that human behavior is produced solely by physiological activity.[88]

In sum, Professor Morse's neuroskepticism holds that, as long as people are agents whose intentions or other mental states are mixed into the chains of causes that produce their behavior, they can be deemed legally responsible and deserving of punishment, even if their mental states and, hence, behavior are the inevitable results of ordinary physical laws.[89] And, he says, nothing in neuroscience or any other science requires

---

libertarians"). For example, Oliver Wendell Holmes wrote that "[t]he reason for requiring . . . [a willed] act is, that an act implies a choice, and that it is felt to be impolitic and unjust to make a man answerable for harm, unless he might have chosen otherwise." HOLMES, *supra* note 13, at 54. Several centuries earlier, Sir Matthew Hale (an early exponent of the "voluntary act" requirement) noted that "ordinarily things naturally act according to the laws and rules and implanted in natural causes," but that "things voluntary [i.e., people] act according to the liberty of their own freedom." SIR MATTHEW HALE, THE WORKS, MORAL AND RELIGIOUS, OF SIR MATTHEW HALE, KNT.: THE WHOLE NOW FIRST COLLECTED AND REVISED. TO WHICH ARE PREFIXED HIS LIFE AND DEATH, Volume I, 369–70 (R. Wilk ed., 1805). *See* Hale's early recognition of the requirement of the voluntary-act requirement in SIR MATTHEW HALE, A HISTORY OF PLEAS OF THE CROWN 412 (Payne ed. 1800). And wonders why the complex of mens rea requirements would ever have developed if judges did not think that the various culpable mental states were within the perpetrator's free will.

86. *See id.*

87. Nichols, *supra* note 77, at 1403 ("Determinism is consistent with the idea that behavior is produced (i.e., determined) by conscious psychological processes").

88. *See infra* Part II.B; *see also* Roskies, *supra* note 28, at 422 (noting that it is not determinism but "reductionism and its attendant consequences, epiphenomenalism or eliminativism, that are most to be feared as a threat to freedom").

89. *Inevitable Mind*, *supra* note 3, at 46.

us to discard the "folk psychology" belief that human beings are agents whose bodily movements and behavior are caused by intentions and reasons, i.e., by mental states, and not (just) by electrochemical operations of brain.[90]

B. Victims of Neuronal Circumstances

Perhaps the number-one bugbear animating Professor Morse and other neuroskeptics is one or another version of the notion that wrongdoers are "victims of neuronal circumstances" (VNC)[91] and, therefore, cannot be justly held responsible for the harms they cause. As Professor Morse tersely says: "Brains do not commit crimes; people commit crimes."[92] The brain is not the responsible party when persons do wrong, and the wrongdoer is not a "victim" of her brain.

The VNC hypothesis originated in a much-noted article by Joshua Greene and Jonathon Cohen.[93] In that article, Greene and Cohen pointed to the considerable evidence that everything a person does is determined by biomechanical neuronal processes and argued that there is, therefore, "something fishy about our ordinary conceptions of human action and responsibility." Accordingly, they said, the legal principles devised to reflect current conceptions of human action "may be flawed."[94] If wrongdoers are (like everyone else) haplessly acting out the prescribed script of fate, how can punishment, even if sometimes useful,[95] ever be morally deserved?

Following the logic of Greene and Cohen, the harm-producing forces of cause and effect can be seen to pass through the wrongdoer's sensory, neural and motor systems in much the same way that, on a simpler level, the electrical force of lightning can pass through a tree, shattering a branch which then falls on a person sheltering beneath. No one would blame the tree just because a causal chain of destructive forces happened to pass through it. And so, one may ask, why should we blame a person who is unlucky enough to become a conduit for a causal chain of harm-producing

90. *Inevitable Mind*, *supra* note 3, at 32–33; *Translation*, *supra* note 3, at 536, quoted *infra* text accompanying note 110.

91. *Inevitable Mind*, *supra* note 3, at 45–46.

92. *Brain Overclaim*, *supra* note 46, at 397. Taken literally, "brains do not commit crimes" is an example of a so-called mereological fallacy (falsely attributing characteristics of an entity to one or more of the parts). If that is all Professor Morse means say, no one could disagree. The statement as a whole is as true (and banal) as saying: "Stomachs don't eat lunch. People eat lunch." But it is not suggested in this article (or, I think, by neuroscientists) that the brain alone causes behavior, and the metonymic use of the word "brain" should not be taken to imply otherwise.

93. Greene & Cohen, *supra* note 8.

94. *Id. at* 1775.

95. Greene and Cohen advocate that punishment be justified on utilitarian grounds. *Id.*

forces that originated elsewhere? To be sure, we are not used to thinking of human bodies as conduits for chains of cause and effect. But, unless harmful physical forces can originate uncaused within the human frame,[96] the complex adaptive input-output system that we call a person is continually susceptible to conscription as a conduit in harm-producing causal chains that originated elsewhere. The paths of causation through brain and body are obviously far more intricate than those involved in a lightning bolt passing through a tree, but the principle is the same. Everything that occurs within a person's skin, and therefore every bodily movement we make, is dependent on events that have previously occurred somewhere else. A person whose neuronal pathways have been commandeered as a vehicle for harmful forces originating elsewhere is a "victim of neuronal circumstances."

Professor Morse is uncompromising in his rejection of the VNC hypothesis, stating that it "completely contradicts common sense and the entirety of our experience."[97] "If VNC is true," he says, "then compatibilism is false" and "no responsibility is possible."[98] The reason VNC would make responsibility impossible is that VNC excludes agency and "[i]t is an incoherent notion to have genuine responsibility without agency,"[99] viz. the ability to act intentionally, rationally and for reasons.[100] It is not entirely clear, however, exactly why Professor Morse considers VNC to be in itself a barrier to attributing responsibility and justifying punishment.

Although the VNC hypothesis is inherently deterministic in nature, the problem that Professor Morse has with VNC is apparently not its determinism. After all, like all compatibilist doctrines, his brand of neuroskepticism does not deny that determinism is true.[101] But this seems to lead Professor Morse smack into another contradiction: If determinism *is* true, then presumably every person's intentions, reasons and other mental states are just as much predetermined by prior events as a VNC's brain states are. Accordingly, even if (as per mental causation) a person's behavior is caused by her mental states rather than her physical brain, she still would be a victim of circumstance if she falls into wrongdoing. The only difference is that she would be, not a victim of *neuronal*

96. And Professor Morse agrees that it probably cannot, *see supra* note 66, and any such uncaused, self-originated forces would, of course, be a violation of the physical law of conservation.

97. *Inevitable Mind*, *supra* note 3, at 46.

98. *Common Sense*, *supra* note 46, at 67; *Inevitable Mind*, *supra* note 3, at 46.

99. *Common Sense*, *supra* note 46, at 67. *Inevitable Mind*, *supra* note 3, at 46.

100. *See Inevitable Mind*, *supra* note 3, at 33; *supra* note 52 and accompanying text.

101. *See supra* notes 61–63. By definition, compatibilism means that responsibility is compatible with determinism.

circumstances, but a victim of *mental* circumstances ("VMC"). And so the question becomes: On what basis could a neuroskeptic think it is perfectly fine to attribute responsibility and blame to VMCs but not to VNCs? In deciding who deserves punishment, what difference does it make that a defendant's wrongful conduct was ineluctably determined by her causally efficacious mental states rather than by her neuronal activity alone? It appears to be a flat-out contradiction to say that VMC transgressors are punishable while mere VNCs are not, a contradiction that Professor Morse does not try to resolve.

C. Burden of Proof

Despite the key role that the mental causation hypothesis performs in folk psychology and the neuroskepticism position, little evidentiary support is offered for the contention that mental causation actually occurs.[102] Professor Morse, for instance, "simply accept[s] the folk psychological view that mental states [have] a genuinely causal role in explaining human behavior."[103] Instead of making an effort to substantiate this "causal role," he preemptively assigns the burden of persuasion to those—such as the advocates of VNC—who would contend that mental causation does *not* occur.[104] What is more, he sets the bar exceedingly high, saying that the burden to disprove mental causation is "enormous"[105] and that the folk-psychology model is justified until "science *conclusively* demonstrates that human beings cannot be guided by reasons and that mental states play *no* role in explaining behavior."[106]

---

102. *See infra* Part IV.

103. *Translation*, *supra* note 3, at 532, n.5.

104. *Common Sense*, *supra* note 46, at 67. ("The burden of persuasion is firmly on proponents of the radical [neuroscience] view"); *Translation*, *supra* note 3, at 553 ("Surely the burden of persuasion is on those who argue [that] the VNC is true").

105. *Neurolaw*, *supra* note 46, at 14.

106. *Free Will*, *supra* note 46, at 251 (emphasis added); *see also id.* at 253. Professor Morse peppers his papers with reminders that (in his view) the burden in on neuroscience to disprove mental causation. *See, e.g.*, *Translation*, *supra* note 3, at 536; *Neurolaw*, *supra* note 46, at 14; *Inevitable Mind*, *supra* note 3, at 46.

This burden-of-persuasion argument would (if accepted) surely be Professor Morse's strongest and, indeed, would effectively decide the debate. Not only has he couched the proposition to be proved in terms of a negative, notoriously hard to prove,[107] he has set the bar unreachably high: Under one-sided ground rules like these, even an overwhelming balance of evidence running against the mental causation hypothesis would have to be disregarded so long as it did not quite "*conclusively* demonstrate … that mental states play *no* role."[108] "The law will be fundamentally challenged only if neuroscience or any other science can *conclusively* demonstrate that the law's psychology is wrong and we are not the type of creatures for whom mental states are causally effective."[109] And, he notes, "neither neuroscience nor any other science has demonstrated that mental states play no independent and partial causal role."[110]

Professor Morse makes no particular effort to justify his bold rhetorical step of assigning the burden to his opponents and demanding conclusive proof. All we get are hints of a kind of "moral inertia" argument; for example, that "we remain entitled to presume that conscious intentions are causal until the burden is met."[111] The problem is that he does not say *why* we are "entitled" to this presumption and, as will be discussed in Part IV, the present evidence for mental causation does not even support a valid *inference* that it exists, much less a presumption.[112]

---

107. *See Folk Psychology*, *supra* note 3, at 4, n.5. Even if every instance of human behavior ever examined by neuroscience were definitively proved to be due solely to physical brain causation, it still could not be ruled out that, at some point in the future, a mentally-caused act might be discovered. This is simply a standard feature of the classic problem of inductive reasoning: It can never provide absolute certainty because we have no right to assume, in Hume's words, that "instances of which we have had no experience must resemble those of which we have had experience." DAVID HUME, TREATISE OF HUMAN NATURE 123 (1826). While inductive reasoning can provide high statistical probabilities of truth (in this case that mental causation never occurs), it cannot rule out the possibility of future counter-examples. *See* Leah Henderson, *The Problem of Induction*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2018), https://plato.stanford.edu/entries/induction-problem/.

108. *See supra* note 106.

109. *Translation*, *supra* note 3, at 534 (emphasis added).

110. *See, e.g.*, *Translation*, *supra* note 3, at 536.

111. *Id.* at 553; *see also Inevitable Mind*, *supra* note 3, at 30 (quoting Jerry Fodor); *Neurolaw supra* note 46, at 14. The obvious reason for elevated burdens of proof is to "skew errors away from defendants for policy reasons." Ronald J. Allen & Michael S. Pardo*, Relative Plausibility and Its Critics* 7 Northwestern Public Law Research Paper No. 18-16 (2018) (citing In re Winship, 397 U.S. 358, 364 (1970); Addington v. Texas, 411 U.S. 418, 424 (1979)). But this rationale suggests that, on constitutional grounds no less, we should try to "skew errors away from" inflicting punishment, not towards it by placing the heavy burden on those who would inflict.

112. Arguably, the burden of proof in disproving the mental-causation hypothesis should be met simply by showing that it is a false inference that is premised on a logical fallacy, as will be discussed *infra* Part IV, coupled with Eddington's analysis showing that mental causation is virtually impossible consistent with the physical nature of the substances involved, discussed *infra* III.B.

There is of course something to be said for preserving the stability of legal institutions,[113] and this might support a degree of inertia in favor of mental causation. But one may be forgiven for suspecting that Professor Morse carries this idea too far. New harms require their own justifications, and when the question is whether to inflict suffering and hardship it would seem not to be a suitable case for woodenly applying moral rationales that were settled for previous cases to new parties or facts. Beyond that, requirements of conclusive proof are not well known in law—"beyond a reasonable doubt" being as high as the burdens seem to get.[114] The burden of proof that Professor Morse has set with such apodictic aplomb appears, at any rate, to be more a rhetorical device than a seriously supported position.

D. The "Fundamental Psycholegal Error"

Neuroscience is concerned with facts about nature, specifically the workings of the brain and its role in causing behavior. It does not make normative findings concerning matters of 'right' and 'wrong,' and normative conclusions cannot be drawn from purely physical facts.[115] Therefore, as Professor Morse is fond of insisting, the causal explanations provided by neuroscience are not normative excuses,[116] and he calls it a

---

113. *See* Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U. CHI. L. REV. 1175, 1178 (1989) ("Law . . . unlike science, is concerned not only with getting the result right but also with stability, to which it frequently will sacrifice substantive justice"); *see also* Holy Props. Ltd., v. Kenneth Cole Prods., Inc., 661 N.E.2d 694 (N.Y. 1995); Estate of Thomson v. Wade, 509 N.E.2d 309 (N.Y. 1987) ("[W]here it can reasonably be assumed that settled rules are necessary and necessarily relied upon, stability and adherence to precedent are generally more important than a better or even a 'correct' rule of law").

114. Gregg Caruso discusses various possibilities and favors a high epistemic bar to prove that free will *does* exist in Gregg D. Caruso, *Justice Without Retribution: An Epistemic Argument Against Retributive Criminal Punishment*, NEUROETHICS (2018), https://link.springer.com/article/10.1007%2Fs12152-018-9357-8. Adam Kolber has discussed a standard that requires proof of key justificatory facts beyond a reasonable doubt in order to justify punishment. Adam J. Kolber, *Punishment and Moral Risk*, 2018 U. ILL. L. REV. 487, 487 (2018) (proposed and defended by Nathan Hanna in *Retributivism Revisited*, 167 PHIL. STUD. 473, 477–78 (2014). With particular reference to the putative justificatory "fact" of free will, Michael Corrado appears to take the position that there may be enough evidence for free will to support some everyday-life decisionmaking but not enough to support brutality as a state response to the problem of criminality. *See* Michael Corrado, *Free Will Fallibilism and the 'Two Standpoints' Account of Freedom*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2943442; Michael Louis Corrado, *Punishment and The Burden Of Proof*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2997654. In an earlier paper, I have argued that the facts supporting a claim of justification have to be shown with "reasonable certainty—the degree of certainty that 'ordinarily prudent' people require as a basis for action generally." John Humbach, *Free Will Ideology: Experiments, Evolution and Virtue Ethics*, Pace Law Faculty Publications 5 (2010), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578445.

115. HUME, A TREATISE, *supra* note 107, at 469. As, noted in a previous footnote, *supra* note 23, nothing in the neuroscience explanation of human behavior denies the important moral distinction between right and wrong.

116. *See Overclaim*, *supra* note 46, at 399; *Inevitable Mind*, *supra* note 3, at 47; *Common*

"fundamental psycholegal error" to think they are.[117] What this truth overlooks, however, is that it is not possible to draw valid normative judgments about conduct or about the moral rightness of punishment without applying normative principles in the light of the actual facts. To the extent that particular facts are relevant to such judgments, the scientific evidence that establishes or disconfirms those facts is also relevant.[118] "Evidence that challenges fundamental legal assumptions can alter the law and its interpretation,"[119] and normative reasoning premised on out-of-date or faulty factual assumptions is unlikely to provide sound results. Therefore, when scientific research uncovers norm-relevant facts that were not previously known, or throws doubt on suppositions previously thought to be true, it will probably be necessary to change the factual premises to which our normative principles are applied.[120] If normative analysis does not use the best factual understandings available, its conclusions will surely go awry.

For example, Professor Morse contends that offenders deserve hardship and deprivation because they are, as persons, considered to be "agents" able to act intentionally and for reasons.[121] But no matter how sound this "agency" justification for punishment may be as a normative matter (a point I will not debate for the moment), it cannot be treated as applicable if its factual predicates—agency and mental causation—are false. Or, to take another example, even though causal explanations are not in themselves excuses, they still can be *relevant* to excuses if they negate the presence of a key element of the alleged offense,[122] such as specific intent[123] or the usual requirement of a voluntary act.[124]

Thus, while it may be tempting to cast aside the evidence of neuroscience on the ground that it tells us only about the "causes" of behavior and not about norms of 'right' and 'wrong,' this argument misses the point. To the extent that justifications for current criminal justice

---

Sense, *supra* note 46, at 55.

117. *Id.* Apparently, although causal explanations are not exculpatory ("excuses"), Professor Morse is happy enough to treat them as *inculpatory*. Specifically, he appears to fully accept the idea that people can be held morally responsible and punished based on the fact their *mental states*, such as intentions, caused their wrongful conduct. Why mental-state causal explanations of behavior are relevant to culpability and brain state explanations are not is a contradiction that is never explained.

118. *See* Kolber, *NeuroLaw Revolution?*, *supra* note 48, at 822 ("[A]nalogizing the relevance of facts about chemicals to norms regulating the handling of the substances).

119. *Id.*

120. *See* Jones & Wagner, *supra* note 8, at 10 (but does not mention challenging the assumption of mental causation).

121. *See supra* notes 52–55.

122. *See Common Sense*, *supra* note 46, at 53.

123. United States v. Veach, 455 F.3d 628 (6th Cir. 2006); MODEL PENAL CODE 2.01(1).

124. *See supra* notes 12–13.

practices are premised on beliefs in mental causation, the facticity of mental causation is not just relevant but crucial to those justifications. By showing that human behavior can be best explained by entirely physical causes, without any need to suppose the agency and mental causes that are essential to moral and (under many current laws) legal culpability, a key justification for inflicting human hardship and deprivation will be lost.[125] Even Professor Morse appears to concede this.[126]

III. THE NEUROSCIENCE THREAT

According to the folk psychology implicit in criminal law, mental states are "fundamental to a full causal explanation and understanding of human action."[127] Based on the findings of neuroscience to date, however, it has already become increasingly doubtful that mental causation is even possible let alone fundamental to an explanation of why people do what they do. To the extent that mental-causation beliefs are the basis for the moral legitimacy of blame, ascribing responsibility and inflicting punishment, neuroscience threatens these practices with a failure of their fundamental premise. By providing a credible and well-documented alternative to the mental-causation hypothesis, the findings of neuroscience strike directly at the factual presuppositions underlying the "guilty mind" justifications for inflicting hardship and deprivation on millions of Americans. As such, they appear to require a fundamental rethinking of the criminal law and justice practices.[128]

---

125. *See* Roskies, *supra* note 28, at 419 ("[D]eterminism is a threat to moral responsibility when the causes of behavior are perceived to bypass mental states").

126. *See, e.g.*, *Inevitable Mind*, *supra* note 3, at 46 ("It is an incoherent notion to have genuine responsibility without agency," i.e., the ability to act based on intentions and reasons"). Obviously, society would continue to need to take coercive measures against dangerous individuals in the interest of public safety. But insofar as the mental-causation hypothesis is the basis for holding that offenders "deserve" hardship and deprivation, the entire approach to criminal justice may need to be revised. The nature of the coercive treatment would likely be entirely different (and vastly less inhumane) if it were seen as merely a necessary and regrettable expedient, rather than a meting out of deserved adverse consequences.

So while facilities for confinement would surely still exist, for incapacitation and as places for rehabilitation. But people would not be sent to these facilities because they "deserve" it but rather because the dangers and risks of the alternatives are simply too socially intolerable to allow. The mission of these facilities would be crime-prevention, not "punishment"; they would be used as a last resort not a default response; and most importantly, they would be designed to actively minimize rather than accentuate the inevitable hardship and deprivation that comes with loss of freedom imposed for the protection of others. *See infra* Part VI.

127. *Free Will*, *supra* note 46, at 255; *see also Common Sense*, *supra* note 46, at 50; *Translation*, *supra* note 3, at 530.

128. *But cf. Translation*, *supra* note 3, at 532–34. Just to be clear again, I make no suggestion that there would (or should) be an end to legal consequences for crime. *See supra* note 126 and *infra* Part VI.

Neuroscience has created serious doubts about the mental-causation hypothesis in a number of ways; three will be elaborated on here. First, modern neuroscience research provides a documented and robust competitor to folk-psychology's mental causation conjecture, which therefore can no longer prevail simply by default. Second, despite extensive neuroscience investigations that would be expected to have turned up some sign of mental causation if it exists, or at least a gap into which it could fit, no such sign or gap has been observed. Thirdly, if mental states like intentions, reasons and beliefs are actually able to cause the body to move, it would flatly contradict the "physicalism" thesis that underlies our general understanding of everything else we know to exist in the natural Universe. These will be considered in turn.

A. Neuroscience Is a Documented Alternative to the Folk-Psychology Conjecture

For thousands of years, the mental-causation hypothesis had no credible competitors.[129] People in the pre-scientific world had no clue (and certainly no evidence) that bodily movements could be produced by neuronal activity alone. They therefore had no choice but to suppose that their mental intentions caused their movements and behavior or else have no causal theory at all. Today's accumulation of evidence detailing the mechanics of neuronal causation threatens the mental-causation hypothesis by putting an end to its explanational monopoly. The old folk-psychology conjectures can no longer prevail simply by default.

As already noted, Professor Morse has attempted to parry this threat by preemptively assigning the burden of persuasion to those who contend that mental causation does *not* occur,[130] requiring "science [to] conclusively demonstrate . . . that mental states play *no* role in explaining behavior."[131] And, he points out, "neither neuroscience nor any other

129. There were, to be sure, attempts to suggest competing theories, such as Malebranche's "occasionalism" according to which God creates successive states of the world from moment to moment and the differences between these successive states gives an illusion of causation, including mental causation. Tad M. Schmaltz, *Nicolas Malebranche*, *in* A COMPANION TO EARLY MODERN PHILOSOPHY 152, 161 (Steven Nadler, ed. 2007); STANFORD ENCYCLOPEDIA OF PHILOSOPHY, *Nicolas Malebranche* § 4 (2013), *https://plato.stanford.edu/entries/malebranche/#Occ* ("God . . . brings it about that our sensations and volitions are correlated with motions in our body"). Leibnitz provided another competing explanation arguing that "the mind and the body are in a 'pre-established harmony,' rather like the clocks that were synchronized by the shopkeeper in the morning, with God having started off our minds and bodies in a harmonious relationship." *See* JAEGWON KIM, PHILOSOPHY OF MIND 171 (2011); *Gottfried Wilhelm Leibniz* § 4.4 (2013), https://plato.stanford.edu/entries/leibniz/#PreEstHa. But these alternative explanations never gained much traction.

130. *See supra* Part II.C.

131. *Free Will*, *supra* note 46, at 251 (emphasis added); *see also id.* at 253. Professor Morse

science has demonstrated that mental states play no independent and partial causal role."[132] The threat that neuroscience poses cannot, however, be so easily dismissed. An explanation cannot be ruled out just because it does not rule out its rival explanation. When people are faced with competing explanatory possibilities, whether in science or in law, the outcome should not depend (if we care about truth) on one side's pre-emptive assignment of an overwhelming burden of proof to the other. It should depend, rather, on which of the alternatives best lines up with the relevant evidence and data. The question is almost never which alternative *perfectly* fits the evidence (which often will be neither), but which one makes the *best* fit.[133] But each remains a possibility, tugging at the honest mind until there is an evenhanded comparative evaluation of the evidence supporting both. The question of which inference provides the best explanation of human behavior, mental causation or causation by the brain, will be considered *infra* Part V.

B. No Sign of Mental Causation in Experimental Findings

If mental causation actually has a role in producing behavior, the growing accumulation of experimental findings[134] would be expected to reveal some sign of it or, at least, some "gap" in neuroscience's causal explanations that mental causation could fill. Yet no sign of any such causation by the mind (or of any other form of psychokinesis) has ever turned up, and mental causation is quite unnecessary to a complete neuroscience picture of the mechanics of behavior production.[135] Not only are there no experimental results that show mental states modifying or affecting patterns of neuronal activity or triggering motor impulses to the body, there is nothing to show how they even could.[136] On the contrary, what the evidence shows is that human beings and all other organisms with brains can produce coordinated, situation-responsive bodily movements and, hence, behavior, by means of neuronal activity alone. To be sure, mere "absence of evidence is not evidence of absence," as the

---

peppers his papers with reminders that (in his view) the burden in on neuroscience to disprove mental causation. *See, e.g.*, *Translation*, *supra* note 3, at 536; *Neurolaw*, *supra* note 46, at 14; *Inevitable Mind*, *supra* note 3, at 46.

132. *Translation*, *supra* note 3, at 536.

133. *See infra* Part V.

134. *See supra* note 4 (1.79 million published articles from 2009 to 2013 alone).

135. David Papineau has called this general line of reasoning the "argument from physiology." David Papineau, *The Rise of Physicalism*, *in* PHYSICALISM AND ITS DISCONTENTS (Carl Gillett & Barry M. Loewer eds., 2001), https://www.academia.edu/819823/The_Rise_of_Physicalism.

136. Indeed, it is a seriously open question whether most organisms, though often capable of complex behavioral repertoires in response to wide variations in their environments, even have minds or "mental states" as understood to exist in human beings.

saying goes. But as the experimental data continues to mount up with no sign of mental causation, this "dog that did not bark"[137] creates an increasingly powerful implication that mental states have no causal role in generating bodily movements or behavior.

One key problem that neuroscience raises for the mental-causation hypothesis is that it seems to rule out, as impossibly causally intrusive, the existence of a mind-brain interface through which the "intentional" contents of mental states could be communicated to the muscle-activating motor neurons. Neuroskeptics do not seem to deny that coordinated contractions of striated muscles (and hence behavior) can only be triggered by electrical signals arriving through motor neurons connecting to the brain and central nervous system. So if there *were* a mind-brain interface that inserts mental decisions into the neuronal processes creating motor signals, how might it work?

This is actually not a new question: As physicist Arthur Eddington wrote as early as 1927,[138] a mental decision can cause bodily movement only if, somewhere in the brain, "the course of behavior of certain atoms or elements of the physical world is directly determined for them by the mental decision."[139] He considered two main classes of possibilities for this intervention: It could occur if the mind could have effects on one or a small number of atoms and "serve as a switch to deflect the material world from one course to another."[140] Or it could occur if the mind could systematically affect a "large groups of atoms"[141] that operate as the components of neurons, as entire neurons or even as larger brain structures. Neither possibility, he concluded, could likely be compatible with physical law.

Eddington noted that the first possible mode of mental-state intervention (individual-atom effects) could at least theoretically occur without obvious contradiction of physical laws. This is because the apparent determinism of cause and effect that we see at the macro level does not apply at the sub-microscopic "quantum" level of atoms and subatomic particles, where apparently uncaused events are

137. Mike Skotnicki, *"The Dog that Didn't Bark:" What We Can Learn from Sir Arthur Conan Doyle About Using the Absence of Expected Facts*, BRIEFLY WRITING (July 25, 2012), https://brieflywriting.com/2012/07/25/the-dog-that-didnt-bark-what-we-can-learn-from-sir-arthur-conan-doyle-about-using-the-absence-of-expected-facts/

138. ARTHUR EDDINGTON, THE NATURE OF THE PHYSICAL WORLD 281 (Cambridge Univ. Press 1928), http://henry.pha.jhu.edu/Eddington.2008.pdf. Eddington wrote in terms of causation by "volitions," but what we are calling *mental* causation is evidently what he had in view.

139. *Id.* at 312.

140. *Id.* at 312.

141. *Id.* at 312–13.

commonplace.[142] Since events at the quantum level can only be foreseen in terms statistical probabilities anyway, a mental decision could conceivably intervene to "deflect" physical events and not be noticed because the mind-body interaction would not, in the given instance, be experimentally distinguishable from random quantum variation.[143] As Eddington went on to point out, however, from a physics standpoint the precise balancing of forces required for this "key-atoms" model of brain operation would be "impossible to maintain" because of the inherent instabilities that result from "physical influences of temperature and promiscuous collision."[144]

The alternative possibility for mental-state intervention (mental decisions "systematically affect large groups of atoms") has major difficulties of its own:

> It is one thing to allow the mind to direct an atom between two courses neither of which would be improbable for an inorganic atom; it is another thing to allow it to direct a crowd of atoms into a configuration which . . . physics would set aside as 'too improbable.'[145]

He said it would be "like the improbability of the atoms finding themselves by chance all in one half of the vessel."[146] This improbability, which is vanishingly close to zero, also represents the likelihood that the mental causation hypothesis is true.

It is notable, perhaps, that Eddington himself was no crusader against the mental causation hypothesis. He clearly *wanted* to believe that there is a causally active place for human "volition" in nature. He argued, for example, that it might be possible to break the vast improbability of mind-body interaction (which he had demonstrated) by considering the "broader unifying trends in the mind-stuff," trends that could lead to "collective behavior" of atoms in response to impinging mental decisions.[147] But Eddington offered no real evidence that any such process could occur and, in the end, his analysis of the mind-brain interface is about as close as science gets to ruling out, once and for all, the possibility of mental-state causation. And it seems very close indeed. Based on the Eddington's

---

142. *Id.* at 299–311.

143. Indeed, Eddington even proposed that, perhaps, the natural probabilities themselves could be modified by mental decisions. *Id.* at 314.

144. *Id.* at 313.

145. EDDINGTON, *supra* note 138, at 314.

146. *Id.*

147. *Id.* at 315. Eddington's hoped-for alternative sounds essentially like a repackaging of traditional beliefs that the soul can affect the course physical events, with "mind-stuff" standing in for the more traditional "soul." Descartes, for example, regarded the mind and the soul as more or less the same thing. Justin Skirry, *René Descartes: The Mind-Body Distinction*, INTERNET ENCYCLOPEDIA OF PHILOSOPHY, https://www.iep.utm.edu/descmind/

analysis there is no *plausible* way, even in theory, that the content of mental states could affect neuronal processes. Rather, mental states are essentially Platonic[148] cave-shadows cast by the physical, no more able to direct the course human conduct than the shadow of a bat can knock a ball across a field.

C. Mental Causation Contradicts the Physicalist View of the Universe

Another way that the findings of neuroscience threaten the belief in mental causation is by highlighting its apparent inconsistency with the precepts of *physicalism*, the thesis that (apart from abstract concepts) "everything is physical"[149] and, in particular, its corollary that every physical event, if it has a cause, must have a physical cause.[150] Physicalism is, in other words, the view of the Universe that rejects the idea that physical objects can be moved around or modified by non-physical forces, such as ghosts, spirits or immaterial minds.

---

148. Plato, REPUBLIC VII 514a (allegory of the cave).

149. *See generally* Daniel Stoljar, *Physicalism,* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2015), https://plato.stanford.edu/entries/physicalism/#TokTypPhy. The basic premise of physicalism is that all of the concrete stuff of universe consists of physical matter and energy, and of nothing else. *Id.* Jaegwon Kim describes physicalism as "the view that there are no concrete substances, in the space-time world other than material particles and their aggregates." Jaegwon Kim, PHILOSOPHY OF THE MIND 211 (1998) (referring specifically to "ontological physicalism" or, as he prefers to call it, "substance physicalism").

For what it is worth, while I generally accept the truth of ontological physicalism, I do not see how such a commitment can be unqualified as long as the nature of consciousness remains unknown. *See* following footnote and the companion to this article, The Neuroskeptic's Physicalism Dilemma, 12 WASH. U. JUR. REV. ___ (forthcoming) [hereinafter Humbach, *Physicalism Dilemma*].

150. This latter is sometimes referred to as the "causal closure principle." *See* KIM, PHYSICALISM, *supra* note 15, at 15. Notice that neither physicalism per se nor the causal closure principle holds that every physical event has a cause, but we generally assume that (except at the quantum level) they do. Notice, too, that while uncertainties about the nature of consciousness may require reservations in our commitment to physicalism, *see* previous footnote, it by no means requires reservations concerning the "causal closure principle." It is, for example, perfectly coherent to think that conscious mental states have a non-physical nature (as it would be, for example, if they are non-physical properties of physical events in the brain—so-called *property dualism*) without also supposing them to have causal efficacy in the physical domain. It is perfectly possible, in other words, that mental states are non-physical but *epiphenomenal*—i.e., causally inert. For more on "property dualism," discussed *infra* note 156 and PHYSICALISM DILEMMA, *supra* note 149.

There is not, of course, conclusive proof that the causal closure principle is true—though everything we know about the Universe (apart from possible mental causation) implies that it is. And one cannot in any case "disprove" mental causation merely by asserting that it is inconsistent with causal closure without blatantly begging the question. What one can do, however, is question whether it is credible that a Universe whose ontology is otherwise so consistently physical throughout would include, in one just-evolved species on one little planet orbiting one very minor star, there would be such a glaring exception to an otherwise universal law that spirits, phantoms and other immaterial forces cannot effect changes in physical reality. There are many things unique and special about the Earth, humanity and terrestrial life but an ability to violate physical law is not known to be among them.

The neuroscience explanation of behavior (human and otherwise) is uncompromisingly *physicalist* in character. It provides a detailed and evidence-based explanation, based on ordinary physical principles, of how neuronal activity can determine and produce bodily movements (behavior), not just in human beings but in every kind of organism having a brain.[151] In contrast to the mental causation hypothesis, which argues that human behavior is produced by a process that is appears to be literally unique in nature, the neuroscience explanation of behavior treats the things that people do as an ordinary part of the physical order.

While believers in mental causation must concede it is an utter mystery how intentions, reasons and other mental states could possibly cause the muscular contractions that bring about movement,[152] neuroscientists see behavior causation as a matter of ordinary physical forces, the kind that are studied in high-school physics: It is a commonplace that channeled currents of electrical forces can carry and process information as well as produce mechanical motion, something that we see all around us every day as the basis for just about everything that makes our modern lifestyles possible, from automatic dishwashers to the control systems of space rockets. Relying on these electrical and electromotive causes and effects, the same kinds of causal interactions that are practically ubiquitous in today's world, biochemically-generated electric currents carry information and process it computationally as the currents are channeled within the neural networks of the brain's connectome by differently-strengthed (potentiated) neural synapses,[153] thereby producing and organizing the motor impulses that cause people and other organisms to do what they do.[154] As far as neuroscience is

151. *See* SAPOLSKY, *supra*, note 6, at 21–77, 535–36 (describing and summarizing how the brain chooses and produces bodily movements); *see also* PASSINGHAM, *supra* note 4, at 66–81.

152. *See Inevitable Mind*, *supra* note 3, at 33 ("[W]e do not have a clue"); *Moore on the Mind, supra* note 46, at 4 ("The brain enables the mind and action, but we have no idea how"); *LawSci*, *supra* note 37, at 24. Professor Morse writes that it should be the "task of neuroscience" to explain how mental states cause action ("agency"), "not to explain it away reductively." *Common Sense*, *supra* note 46, at 47. But this charge to neuroscience begs the question of *whether* mental states cause action. It is the task of neuroscience to explain how human behavior is produced, but not to show how a particular pre-chosen cause of behavior, favored on policy, religious or grounds might be able to produces it.

153. *See generally* KOLB & WINSHAW, *supra* note 6, at 109–57, 164–70; PASSINGHAM, *supra* note 4, at 66–81; SAPOLSKY, *supra*, note 6, at 21–77, 535–36 (describing and summarizing how the brain chooses and produces bodily movements); and other sources cited *supra* note 6.

154. It is not, to be sure, a simple matter to map the connectome or study how differently-strengthed synapses channel electrical forces through its neural networks. But no one asserts, I think, that sequences of synaptic firings in the brain are willy-nilly or denies that there *are* traceable "pathways" which nervous impulses follow through the complex and plastic circuitry of the brain, or that these patterns of excitation run (and are roughly traceable) from the sensory organs through the areas of the pre-frontal cortex (among others) to the motor neurons that activate muscular movements. *See* KOLB & WINSHAW, *supra* note 6, at 109–57, 164–70. It is, however, beside the point that no one can know what a given brain will do at a given moment or in a given situation. It could likewise be

concerned, human behavior, like everything else that happens within the fabric of the Universe, results from interactions among bits of matter and energy in accordance with physical law, not from conscious mental states.

One can quibble that physicalism is not (yet) able to explain everything,[155] but that is hardly a justification for favoring a non-physical causal explanation of events when a physical one is plausible. A bump in the attic at night makes most of us think that something tipped over, not of spirits and ghosts. When mental causation's imaginative contrivances like "property dualism,"[156] emergentism[157] and the irreducibility of mental causes[158] are contrasted with the scientific building blocks of the neuroscience explanation of behavior, it is difficult to see the mental-causation hypothesis as anything but an ontological outlier—an intrinsically non-material explanation of a kind that is generally eschewed in modern mainstream scholarship.[159]

---

said that no one can say what pattern a kaleidoscope will show after fifty turns of the wheel. But that does not give reason to doubt (in either case) that what is going on is a purely physical process dictated by purely physical laws.

155. A notable example is consciousness itself. *See supra* notes 35–36; PHYSICALISM DILEMMA, *supra* note 149.

156. Property dualism is the idea that physical events can have non-physical properties that cannot be fully explained by the properties of the underling physical substrate itself. *See* KIM, PHYSICALISM, *supra* note 15, at 20–22, 156–61; Howard Robinson, *Dualism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY 2.2 (2016), https://plato.stanford.edu/entries/dualism/#ProDua. For example, a painting of a park in Paris may physically consist of nothing but blobs of dried pigment on a stretched cloth; nonetheless, the image can have properties, such as eye-appeal, artistic balance or meaning, that cannot be entirely accounted for solely by reference to the physical laws governing the atoms that comprise the blobs of paint and cloth. Property dualism is posited to avoid the trap of "substance dualism," the now generally discredited theory of Descartes that mind and body are two different *substances*. *See* KIM, PHYSICALISM, *supra* note 15, at 156 ("Applied to the present context, substance dualism posits the existence of both physical bodies and nonphysical minds whereas physicalism admits the existence only of physical objects and processes"). Unlike substance dualism, property dualism accepts that there is only one substance—the physical—but it finds hope for mental causation in the idea that physical substances, such as arrangements of atoms, can have non-physical properties, such as consciousness (just as the painting Paris may have non-physical properties such as beauty and meaning).

157. Emergentism is a form of property dualism which holds that certain arrangements of atoms can give rise to discernible objects and properties whose behavior cannot be entirely accounted for by the properties of the atoms themselves, for example, arrangements of dots can be viewed as a picture that has properties over and above the properties of the dots themselves, or certain arrangements of atoms or neural firings can give rise to conscious states with properties over and above those of the atoms or neural firings themselves. *Id.* at 157–58 ("In addition to physical properties [of the substance of the brain], there are physically irreducible domains of emergent properties, of which mental properties are the leading candidates.").

158. In this context, "irreducible causes" refers to the idea that the causation of events by mental states cannot reduced to underlying physical causes on the theory that mental-state causes are not merely physical in nature but are something that cannot quite be captured in physical terms. *See* Susan Haack, *Brave New World: On Nature, Culture, and the Limits of Reductionism*, in EXPLAINING THE MIND 37 (Bartosz Brozek et al. eds., 2019); PHYSICALISM DILEMMA, *supra* note 149.

159. Notably, although property dualism, emergentism and irreducibility (non-reductive physicalism) all can, to be sure, be useful as ways of characterizing reality for various purposes, none

IV. THE EVIDENCE FOR MENTAL CAUSATION

A. The Inference from Introspection

There is not much to say about the evidence for mental causation because there is so little of it. Indeed, there is no evidence at all as to how intentions or other mental states (as distinct from co-occurring brain activity) could inject themselves into chains of physical cause and effect to produce bodily movements.[160] Not only is there no evidence as to what the "mind" consists of or how mental states come into being,[161] there is nothing whatever that shows how the immaterial mind (as opposed to a biomechanical brain), could possibly do the things it is supposed to do such as think, classify, compare, recall, reason, direct behavior and so on. Our understanding of the ontological nature of the mind, as opposed to the brain, has advanced little beyond the stage it had reached when (still often called the "soul"[162]) it was pondered by Descartes and Pascal.[163] Rather,

---

of them entails the possibility that mental states have the power to modify the course of physical reality. Their primary usefulness for present purposes is, therefore, only to assert that a physicalist outlook *does not rule out* mental causation. They do nothing to affirmatively establish that it exists in the first place. All are discussed at greater length in PHYSICALISM DILEMMA, *supra* note 149.

160. *See supra* notes 138–48. The parenthetical "as distinct from co-occurring brain activity" denotes that I am setting aside for the moment arguments that mental causation exists in the sense that mental states can be seen to partake of the causal powers of underlying brain states either because the mental states are identical to the brain states or because they are "supervenient" on them. *See* KIM, PHYSICALISM, *supra* note 15, at 93–147 (describing and critiquing such arguments). For one thing, that sort of piggy-back mental causation does not seem to correspond to the law's conception of mental states' distinct role in criminal behavior ("distinct" from the physical elements, or *actus reus*, of crimes). That is, when the law speaks of a person's intentions, knowledge, recklessness or other mental states, it is not referring to something in the person's *physical* makeup or features of her neuronal physiology. Rather, the law's conception of intentions and other mental states refers to states of conscious awareness that are in their essence *not* physical (i.e., *not* fully "reducible" to purely physical components such as the physical neuronal networks of the brain and body). *See Inevitable Mind*, *supra* note 3, at 34.

A more important reason for setting aside piggy-back mental causation is that mental states whose effects identically mirror the effects of physical states on which they depend would seem to have no distinctive moral significance for purposes of moral responsibility. Even if such mental states could be properly said to have causal effect, they would seem to add nothing to the causal picture. If the underlying physical-causation events alone do not suffice to justify ascribing responsibility and inflicting punishment, then neither would the causally redundant mental states that shadow them. For a physicalist, moreover, they would be excluded as causes by the exclusion principle, which forbids that sort of misleading double-counting and misattribution of causes. For further discussion of the exclusion principle, see *infra* note 198 and PHYSICALISM DILEMMA, *supra* note 149.

161. *See Inevitable Mind*, *supra* note 3, at 33 ("[W]e do not have a clue") and other authorities cited *supra* note 152. Professor Morse attempts to finesse this absence of evidence by declaring it is the "task of neuroscience" to explain how mental states cause action ("agency"), "not to explain it away reductively." *Common Sense*, *supra* note 46, at 47. But in assigning this task to neuroscience, he begs the question of *whether* mental states cause action. To be sure, it is the task of neuroscience to explain how human behavior is produced, and it does, but there is no reason to think neuroscience is required to show how the *mind* is able to produce it. *See also* notes 35–37.

162. As Descartes wrote: "I consider the mind not as a part of the soul, but as the whole soul

mental causation is largely understood as essentially a type of intra-body psychokinesis, something that just happens when non-physical forces of "intention" and "reasons" emanate from conscious awareness and intervene in ordinary physical interactions, changing the course of events.

Wittgenstein noted this problem of pinning down the causes of bodily movements when he wrote, in a passage that Professor Morse is fond of quoting[164]:

> Let us not forget this: when 'I raise my arm', my arm goes up. And the problem arises: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?[165]

The answer given by neuroscience would be, of course, that what is "left over" are muscle contractions prompted by neuronal activity in the brain and central nervous system (or, more elaborately, by successions of coordinated firings of synapses in populations of neurons that initiate[166] and organize[167] muscle-activating motor signals that are sent through motor neurons to the arm).[168] While Wittgenstein did not mention brain states and neuronal firing as possible causes of bodily movements (he wrote before the era of modern neuroscience), he was clear that he did not believe it was the "will" that caused his arm to rise. "[W]illing." he wrote, "is *merely an experience*," [169] perhaps "kinesthetic sensations,"[170] but not an "instrument" to bring about bodily movement.[171] Despite careful

---

that thinks." Robert Pasnau, *The Mind-Soul Problem*, *in* MIND, COGNITION AND REPRESENTATION: THE TRADITION OF COMMENTARIES ON ARTITOTLE'S DE ANIMA 4 (Bakker & Thijssen, eds. 2007).

163. *See* Florence M. Weinberg, *The Idea of the Soul in Descartes and Pascal*, 8 FRENCH FORUM 5 (1983).

164. *E.g.*, Morse, *Common Sense*, *supra* note 46*,* at 59; Morse, *Neurolaw*, *supra* note 46*,* at 5; Morse, *Free Will*, *supra* note 46*,* at 258; Morse, LawSci, *supra* note 37, at 23.

165. LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS ¶ 621. (G.E.M. Ancombe trans. 1953) [hereinafter WITTGENSTEIN, INVESTIGATIONS]. Wittgenstein's main point seems to have been that, semantically, the difference between 'I raise my arm' and 'my arm goes up' is that the first statement identifies who or what raised Wittgenstein's arm while the second one does not. However, Professor Morse quotes the passage perhaps thinking it is suggestive of the idea that "what is left over" is mental causation, though the overall tenor of Wittgenstein's remarks *in situ* are, if anything, exactly the opposite. *See infra* notes 169–72.

166. Probably in either the prefrontal cortex or the posterior striatum along with the medial parietal cortex, depending on whether the action was novel or habitual, respectively. *See* PASSINGHAM, *supra* note 4, at 67–72.

167. Typically, in the various motor areas (premotor areas, motor cortex and supplementary motor cortex). *Id.*

168. PASSINGHAM, *supra* note 4, at 66–81; SAPOLSKY, *supra*, note 6, at 21–77, 535–36, describing and summarizing how the brain chooses and produces bodily movements; and other sources cited *supra* note 6.

169. WITTGENSTEIN, INVESTIGATIONS, *supra* note 165, at ¶ 611 (emphasis added).

170. *Id.* at ¶ 621.

171. *Id.* at ¶ 614; *see also id.* at ¶ 615. ("Willing, when not meant as a kind of wish, must be *the action itself*. It cannot come to a stop before the action.") Trans by author. In this, Wittgenstein echoes

introspection, Wittgenstein never reported that he ever detected any specific sensation of mental causation itself. Neither, to my knowledge, has anyone else.[172]

It may come as a surprise to some to hear there is so little evidence for mental causation. After all, as Professor Morse writes, "virtually every neurologically intact human being takes for granted . . . the subjective experience of first person agency, the experience of mental causation that my bodily movements and thoughts are caused, roughly speaking, by my intentions."[173] In statements such as this, the suggestion is made that there is a veritable mountain of evidence for mental causation, based on potentially billions of first-person reports. There is, however, a serious problem with this introspective "evidence" of mental causation.

The problem is that, even conceding the reliability of introspection, it is almost surely an overclaim to speak of the "the experience of mental causation" itself. The fact that every "intact human being" has a subjective experience of agency—the feeling of "I did it"—does not mean anyone ever directly experiences or observes the mental-causation event itself. Indeed, as David Hume pointed out, probably irrefutably, no one ever directly experiences or observes *any* kind of causal event in itself: There is *no* kind of causation that we can ever know directly from experience alone.[174] When a causal event occurs, the only events that are experienced in conscious awareness are the events that *bracket* the causal event, not the "power or necessary connexion . . . [or] quality, which binds the effect to the cause."[175] In the case of mental causation, therefore, it is almost certain that all anyone ever directly experiences are the events that happen before, and then after, the supposed mental-cause interaction occurs. To be more specific, the most that "every neurologically intact human being" *actually* experiences is: (1) a consciousness of intentions, reasons or other similar

---

Schopenhauer. I ARTHUR SCHOPENHAUER, THE WORLD AS WILL AND REPRESENTATION 100–02 (E.F.J. trans. 1968) (1859) [hereinafter SCHOPENHAUER, THE WORLD].

172. *See* Moore, *What Is the Sense of Agency*, *supra* note 78, for an illuminating discussion. I am for the moment deferring discussion of the "sense of agency" that often accompanies voluntary bodily movements. *See infra* notes 185–92 and accompanying text.

173. Morse, *Inevitable Mind*, *supra* note 3, at 34–35. *See also* Morse, *Translation*, *supra* note 3, at 551 ("Virtually every neurologically intact person consistently has the experience of first-person agency, the experience that one's intentions flow from one's desires and beliefs and result in action"). Professor Morse also likes to quote Jerry Fodor to the effect that, if we are wrong about these assumptions, "it's the wrongest we've ever been about anything except belief in the supernatural." *See, e.g.*, Morse, *Inevitable Mind*, *supra* note 3, at 30; Morse, *Neurolaw*, *supra* note 46, at 14. The weakness of this sort of *ipse dixit* argument does not require comment—and this is quite apart from the irony that, based on the physical *evidence* we have, it appears that belief in mental causation *is* a belief in the supernatural.

174. DAVID HUME, AN ENQUIRY CONCERNING HUMAN UNDERSTANDING 60 [Sec. VII Pt. I ¶ 6] 1988) (1748); [hereinafter HUME, ENQUIRY]; *see also* Saunders, *supra* note 13, at 454 n.48.

175. HUME, ENQUIRY, *supra* note 174, at 60 [Sec. VII Pt. I ¶ 6].

mental states relating to possible bodily movements in the immediate future, and then (2) a consciousness of bodily movements that occur immediately following such mental states, that seem to correspond to them, and that apparently have no other cause. In other words, while there is direct conscious awareness of experiences that *bracket* the moment when mental causation is supposed to occur, no one ever observes the actual causal interaction itself. And, *a fortiori*, no one ever introspectively observes whether that interaction is an event of mental causation or is, rather, purely a sequence of neuronal events.

The epistemic challenge posed by fact that people never experience a mental-causation interaction directly is not, as already noted,[176] something that is unique to *mental* causation (as opposed to other causation), but the epistemic challenge posed by mental causation goes even deeper. Not only does no one ever observe the "power or necessary connexion" of mental causation itself but no one ever even observes any events in which a mental-causation interaction could be embedded—in the way, for example, that one can observe a billiard ball knock another ball across the table. In the case of the billiard balls, we may never see the ultimately causative subatomic forces at work imparting energy from one ball to the other, but we at least can observe the physical interaction in which those forces could work. We can, in other words, observe the general mechanism of the causation. From this knowledge of mechanism plus the before-and-after facts, we can reasonably infer that the causal event occurred. With mental causation, however, it is as though we see the first billiard ball approach the second, then see the second ball scoot away, but never see or hear the causal event of impact, the event where the causal forces are in operation.[177] It is as if, watching a movie, several critical frames have been cut out between the depiction of the cause and depiction of the effect, so we can only *infer* that there was a connection between the two. Thus, when we experience our own intentions to act, and then our bodily movements, the best we can do is make an *inference* of mental causation based on these two bits of information—the experience of what went before (the intentions) and of what came after (the corresponding movements).[178] Mental causation is a story we tell to connect the two

176. *See supra* 174–75.

177. Professor Morse cites an experiment demonstrating that people are unable to discern whether their actions were caused by their intentions or by something else. *See* Morse, *Translation*, *supra* note 3, at 548 n.31. This experiment provides empirical support for the hypothesis that people do not reliably observe or otherwise experience a causal connection between their intentions and their actions—at least not under the conditions of the experiment.

178. At noted above, I am for the moment deferring discussion of the "sense of agency" that often accompanies voluntary bodily movements. *See infra* notes 185–92 and accompanying text.

percepts, but it is still just a story—not a percept.

There is, however, a big problem with relying on this inference of mental causation that is based solely on what went before and what came after. A case for mental causation that rests on this inference is premised on a logical fallacy. The fallacy is *post hoc ergo propter hoc* (literally, "after this, therefore on account of this").[179] It is logically fallacious to infer that mental causation occurs based solely on the fact that bodily movements are observed to follow conscious mental states that prefigure the movements. The *post hoc* inference is a "false cause"[180] fallacy. It is "inherently mistaken"[181] and not valid.[182]

---

179. Leo Groarke, *Informal Logic*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2017), https://plato.stanford.edu/entries/logic-informal/; Hans Hansen, *Fallacies*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2015), https://plato.stanford.edu/entries/fallacies/.

180. Bradley Dowden, *Fallacies,* INTERNET ENCYCLOPEDIA OF PHILOSOPHY. http://www.iep.utm.edu/fallacy/#FalseCause.

181. Hansen, *supra* note 179. One of the reasons it is inherently mistaken is that, even if there is a regular temporal correlation between the two kinds of events, this correlation may be due other facts that in no way involve a cause-and-effect relationship between them. For example, the correlation might be due to the fact that the two events both have the *same* cause or causes, and such a "common-causation" possibilities are not infrequent. *See* Jaegwon Kim, *Causation and Mental Causation*, *in* CONTEMPORARY DEBATE IN THE PHILOSOPHY OF THE MIND 7 (Brian McLaughlin & Jonathan Cohen eds., 2007) http://www.colbud.hu/bloewer/Kim_Causation_and_Mental_Causation_-_Debates-1.pdf. Indeed, as will emerge in later discussion, the best explanation for the observed temporal relation between mental states and actions is that both the mental states and the actions depend on a third factor, namely, physical brains states.

182. It should be noted that the *post hoc* fallacy problem infecting judgments about mental causations is not the same as (and is far more serious than) the general problem of establishing causation that was famously illuminated by David Hume, who argued, probably irrefutably, that that the existence of causation as such can never be known from experience alone. *See* Saunders, *supra* note 13, at 454 n.48. According to Hume, we can observe correlations and "constant conjunctions" among events, but causation per se is never actually observed. *See e.g.*, HUME, ENQUIRY, *supra* note 174, at 36 [Sec. IV Pt. II, ¶ 3] ("[I]f you insist, that the [causal] inference is made by a chain of reasoning, I desire you to produce that reasoning"). While Hume's insight is important, it is vulnerable to the objection, made by Immanuel Kant, namely, that it leads to a conclusion in which "the very notion of a cause would entirely disappear." IMMANUEL KANT, CRITIQUE OF PURE REASON *at* Introduction, II (1781). For purposes of the discussion at hand, however, it will be assumed that "the very notion of cause" has not disappeared, that events definitely *can* generate or produce the occurrence of other events, and that the question is, rather, whether *in any given instance* there is in fact a causal relationship.

As a practical matter, the way to show whether a causal relationship exists is normally to demonstrate that, in addition a brute correlation, the putatively causal event is comprised of smaller component events (energy flows, momentum transfers and the like) whose causal efficacy is not in dispute and that connect the putatively causal event with the one it supposedly caused. *See* KIM, PHYSICALISM, *supra* note 15, at 47. That is to say, causation is proved by showing an explanatory *mechanism* that is comprised of component causes, and components of components, going all the way down, at least in principle, to the four basic forces of physics acting on the most granular bits of matter. *See* ENCYCLOPEDIA BRITANNICA, *Fundamental Interaction*, https://www.britannica.com/science/fundamental-interaction.

The problem with the case for mental causation is that it is based, not on a demonstration of a mechanism comprised of well-accepted component causes, but rests entirely on a single, unique and ad hoc causal leap based on a single oft-repeated type of unexplained correlation. By contrast, the neuroscience inference that behavior is caused by neuronal activity (brain states) rests, not on

Thus, the apparent mountain of first-person evidence for mental causation from "virtually every neurologically intact human being" is, on further consideration, not so imposing after all. It is in its essence nothing more than billions of replications of the same bit of evidence: namely, the fallacious inference that nearly every person draws from the same unexplained correlation—the fallacious inference, countlessly repeated, that bodily movements are caused by the mental states, such as intentions, that prefigure them. And the same fatal weakness applies to all of the "common-sense first-person and third person evidence"[183] from every "intact" person that supposedly shows, in its multiplicity of replications, that mental causation exists. It is as if (to paraphrase another well-known line of Wittgenstein) someone were to buy billions of copies of the morning paper to assure himself that what it says is true.[184]

If the belief in mental causation is founded on a logical fallacy, one may wonder why it is so much more readily accepted, and harder for people to let go of, than other similarly paranormal dubiosities like psychokinesis, extra-sensory perception and messages that arrive via ouija boards. Part of the reason is, no doubt, the "absence of surprise"[185] and

---

correlation, but on detailed evidence of intricate mechanisms of component causes (neuronal activity) which, themselves, can be further broken down into familiar and well-accepted underlying causes, such as chemical and electrochemical interactions, the physical behavior of ions, and electromotive forces of attraction and repulsion—causes whose efficacy is not a matter of controversy and that is confirmed in a huge variety of different contexts across the Universe. For example, electrically-charged subatomic particles can be observed in countless different contexts to produce motion as they attract and repel, to flow in currents, sometimes carrying larger particles (ions) with them, and to be organizable by various means into computation-capable arrangements and configurations that are able to assemble information inputs and organize coherent patterns of electrical outputs, which can serve as causes of still further events. These kinds of causal interactions are an ordinary, everyday feature of the universal physical order. We know them, not as a special introspective insight, but as a matter of general physics.

In sum, even though every causal explanation must, at some level, make the Humean "leap" that there is such a thing as cause in fact, the leap required by the neuroscience explanation of behavior is of a completely different order from that made by proponents of mental causation: We must make the leap of assuming that what appear to be four basic forces of nature are in fact causal. It is true that we do not know *what causes* electrical charges to attract and repel, and science must indeed make an unprovable Humean *inference* that there is some force (named electromagnetic force) that makes them do it. But what we do know, as well as we can know anything, is that electrically charged particles and ions unfailingly *do* attract and repel wherever they are observed, and that they do so in strict accordance with physical laws, and that, indeed, if they did not, our entire modern electricity-powered civilization would fall asunder. The electrical causes on which neuroscience explanations are fundamentally based are not sui generis or a matter of doubt.

The difference between the fallacious inference of mental causation and the neuroscience explanation of behavior as caused by brain activity is further discussed *infra*, Part V.

183. *Overclaim*, *supra* note 46, at 401–02.

184. *Cf.* INVESTIGATIONS, *supra* note 165, at ¶ 265 ("As if someone were to buy several copies of the morning paper to assure himself that what it said was true").

185. *Id*. at ¶ 628; *see also* Valérian Chambon et al., *From Action Intentions To Action Effects: How Does The Sense Of Agency Come About*?, FRONTIERS IN NEUROSCIENCE 1, (2014),

associated sense of agency[186] that is experienced when voluntary bodily movements occur.[187] The sense of agency is, however, ambiguous with

https://www.frontiersin.org/articles/10.3389/fnhum.2014.00320/full. ("[We] clearly recognize *failures* of agency when we experience actions that do not unfold as expected or fail to produce intended effects. . . [O]ur sense of "authorship" becomes apparent only when it is falsified, resulting in a break of the flow from intentions to action effects that normally characterize experience"). *See also* D. O'Connor, *The Voluntary Act*, 15 MED. SCI. & L. 31, 32 (1975), quoted in Saunders, *supra* note 13, at 457 n.59 ("Ordinary bodily movement is, in one sense, automatic. This means that one has no awareness of the series of events that precede it . . . It is not until there is a breakdown or malfunction that one looks to find causes for abnormal behaviour just as one begins to concern oneself with the working of motor cars only when something goes wrong").

186. *See* Moore, *What Is the Sense of Agency*, *supra* note 78, at 2 (describing the sense of agency as an "elusive experience" that is "phenomenologically thin" and "quite unlike conscious experience in other modalities, especially vision, where our experiences are phenomenologically strong and stable"). Not only is "the sense of agency is not an infallible reproduction of objective reality," *id.* at 2, it appears to be highly context-dependent, sensed as stronger in moral contexts as opposed to non-moral (economic) ones, and stronger with respect to severely negative effects than with moderately negative effects. Moretto G. et al., *Experience of Agency and Sense of Responsibility*, 20 CONSCIOUSNESS AND COGNITION 1847 (2011), https://www.ncbi.nlm.nih.gov/pubmed/21920776. In short, the sense of agency may be a very useful tool for calibrating ourselves in social interactions, but it is not so useful as an accurate measure of the mechanisms of causal interactions.

Indeed, there is apparently some experimental evidence that the "sense of agency" can be artificially triggered in the absence of agency by direct electrical stimulation of certain areas of the brain associated with producing bodily movement. Reportedly, these stimulations sometimes can generate "an irrepressible urge to perform an action, or . . . an anticipation that action is about to occur," which was sometimes followed by actual motor activity, sometimes not. *See* Uri Maoz & Gideon Yaffe, *supra* note 31, at 128 (citing Itzhak Fried et al., *Functional Organization of Human Supplementary Motor Cortex Studied by Electrical Stimulation*, 11 J. NEUROSCIENCE 3656, 3666 (1991); Michel Desmurget et al., *Movement Intention after Parietal Cortex Stimulation in Humans*. 324 SCIENCE 811, 813 (2009)). Stimulation of other brain areas could sometimes generate bodily movements which the patient evidently was unconscious of (or, at any rate, strongly denied. *Id.* These experiments suggest that "the sense of agency may not be as strongly coupled with voluntary movement as humans generally experience them to be, and that—at least under rather abnormal circumstances—humans may experience agency over phantom actions and carry out actions with no accompanying sense of agency." *Id.* Maoz and Yaffe conclude that it is "far from clear" that "these experiments show that bodily movements are not guided by conscious mental activity, as required by the law's voluntary act requirement." *Id.* They do, however, seem to confirm that people do not experience mental causation interaction itself, as evidenced by the fact that the experimental subjects never seemed to mention it but only mentioned experiencing an "urge" to move or "anticipation" of movement along with subjective experiences of movement or non-movement.

See also Chambon et al., *supra* note 185, at 1 ("We rarely have an intense, clear phenomenology of agency, but we clearly recognize failures of agency when we experience actions that do not unfold as expected or fail to produce intended effects.... our sense of "authorship" becomes apparent only when it is falsified, resulting in a break of the flow from intentions to action effects that normally characterize experience").

187. Apparently, the "sense of agency" is due in part to neuronal processes in the angular gyrus of the parietal lobe that monitor and compare the degree of correspondence between the results of brain-generated movements and. their computationally predicted results as well as, perhaps, monitor the "fluency" of computing the choice of actions (the relative "effortlessness" of the choice). *See id.* ("[T]he angular gyrus (AG), a parietal brain region which has been shown to compute retrospective agency by monitoring mismatches between actions and subsequent outcomes . . . may also code for a prospective sense of agency, by monitoring action selection processes in advance of the action itself, and independently of action outcomes"). In other words, the sense of agency is computed in the brain as an *inference* much like the inference of causation said to be experienced in consciousness. The researchers do not mention noticing any evidence that detection by the brain of mental causation (as

respect to the mental-causation question that it is invoked to answer.

It is said that the sense of agency is the "feeling of being in the driving seat when it comes to our actions."[188] Whenever someone feels a sense of agency with respect to a given bodily movement, it surely makes sense for her to infer, at least *prima facie*, that the movement was generated inside her person rather than solely by some external force such as, say, a sudden gust of wind. But it would be a mistake to assume that the having a sense of agency is the same thing as "sensing mental causation."[189] The conscious experience of agency has been shown to be "linked to low-level sensorimotor processes"[190] that compute the sense of agency "by matching predicted and actually experienced consequences of movement."[191] A "large body of evidence suggests that the sense of agency . . . strongly depends on the degree of congruence versus incongruence between predicted and actual sensory outcome . . . ."[192] In other words, the sense of agency appears to be computationally inferred in the brain, not by monitoring mind/body interactions that constitute mental causation, but

---

opposed to its detection of purely physiological events associated with the generation of motor signals and detection of resultant movements) plays any role in the computation of the sense of agency.

188. Moore, *What Is the Sense of Agency*, *supra* note 78, at 1; *see also* Matthis Synofzik et al., *Beyond the Comparator Model: A Multifactorial Two-Step Account of Agency*, 17 CONSCIOUSNESS AND COGNITION 219 (2006) ("[It is] the registration that we are the initiators of our own actions").

189. A distinction Professor Morse glided over. *See Scientific Challenges*, *supra* note 46, at 847.

190. Moore, *What Is the Sense of Agency*, *supra* note 78, at 1.

191. Chambon et al., *supra* note 185, at 1. According to Chambon et al., using the "influential" comparator account," "agency is computed by matching predicted and actually experienced consequences of movement." Thus, the comparator model allows for two specific predictions. First, sense of agency should be strong when there is a close match between the predicted and the actual sensory consequences of an action, and should be reduced when predicted and experienced consequences do not match. Second, sense of agency necessarily occurs late, i.e., after an action has been performed, and sensory evidence about the consequences of action becomes available." *Id.* Chambon et al., also have found evidence of a brain mechanism that can compute the sense of agency *prospectively*, but it too has nothing to do with detecting mental causation per se. Rather, it works by monitoring the relative "fluency" of computing the choice of which action to perform (the relative "effortlessness" of the choice). *Id.* at 6–8.

There is some debate over the adequacy of the comparator model to fully explain the feelings and judgments of agency, but no one suggests that the sense of agency is anything but a computed inference or is based on direct detection of mental causation. *See also* Glenn Carruthers*, The Case for the Comparator Model as an Explanation of the Sense of Agency and Its Breakdowns*, 21 CONSCIOUSNESS & COGNITION 30 (2010); Glenn Carruthers, *A Comparison of Fortunes: The Comparator and Multifactorial Weighting Models of the Sense of Agency,* PROCEEDINGS OF THE 9TH CONF. OF THE AUSTRALIAN SOC'Y FOR COGNITIVE SCIENCE (2010); Nicole David et al., *The "Sense of Agency" and Its Underlying Cognitive and Neural Mechanisms*, 17 CONSCIOUSNESS & COGNITION 523 (2008); Synofzik et al., *supra* note 188, at 3.

192. David et al., *supra* note 191, at 524. The fact that the brain computes the sense of agency merely as an inference and not be by monitoring mental-cause interactions directly is not, of course, proof that mental causation does not exist. But it does show that the sense of agency can hardly be treated as relevant evidence *for* mental causation.

rather on the basis of before-and-after correlations between the pre-motor neuronal preparations and subsequently perceived bodily movements. So even if the sense of agency can be fairly taken to imply that the immediate cause of a bodily movement was something within the person in question, it is a pure interpretational flourish, unsupported by any factual basis, to say that the causal "something" was the person's mental states rather than her brain activity. Accordingly, the sense of agency adds no experiential evidence of mental causation that could bolster the logically fallacious inference based on bare correlation that we already know about.

Another reason why belief in mental causation may be so hard to let go of is that the specific content of intentions and reasons seems to correlate so closely with the bodily movements that follow or co-occur with them. If a person consciously intends to raise her right hand, it is always the right hand that goes up, never the left.[193] While this reliable correlation is impressive, it becomes less impressive when one considers that the brain states (neuronal activity) needed to produce bodily movements would have exactly the same reliable correlation with the movements they produce. That is, a pattern of neuronal activity that raises the right arm today would always also raise the right arm tomorrow.[194] As

193. Hedda Hassel Mørch has made a well-reasoned argument that this coincidence supports an evolutionary argument for mental causation (or, as she calls it, "phenomenal powers") by invoking the methodology of inference to best explanation. *See* Mørch, *The Evolutionary Argument*, *supra* note 9. Her analysis is, however, in terms of the correlation between actions and the qualia of pain and pleasure sensations, not in terms of the correlation with the kinds of mental states (intentions, reasons, etc.) most relevant to criminal law. She makes a strong argument that the qualitative experience of "pain," in particular, makes it highly improbable that it would have evolved to be "what it is like" unless the feeling of pain had the power to prompt action.

In the end, however, I disagree with her conclusion that qualia can be causally effective. The way that things look, smell, feel, taste and sound may greatly enrich our conscious experience and therefore seem of great subjective importance, but as far as anyone knows, the exact qualities of these experiences—what they are "like"—is just as arbitrary as the coloration of isotherms on a newspaper weather map. Even though the feeling of pain is especially poignant, the more obvious physical explanation for its putative causal power is that, whenever we experience pain, a lot of other neuronal activity is going on (some of which we experience as "emotion") as part of the body's procedures to prepare itself to deal with a possible crisis. There is no reason to think that any motor neuron activity that we believe to be produced by the feeling of pain is, in fact, produced by the neuronal activity that gave rise to the pain qualia in the first place (and this would be the preferred interpretation—absent evidence to the contrary—for anyone with physicalist leanings).

194. I do not mean to question "multiple realization" here. That is, I am not saying that a given bodily movement can only be produced by a single specific sequence of neuronal activity and no other. What I am saying, though, is that a given sequence of movement-producing neuronal activity can only produce one specific kind of movement—with which it (and its supervenient mental state) is correlated. It could not cause the right arm to rise on one occasion and the left arm to rise on the next. I cannot say there is a broad evidentiary base of experimental results confirming the close correlation that logically must exist (for a physicalist) between mental states and particular neuronal activity across a broad spectrum of possible mental states, but there is evidence that such a correlation exists and none, to my knowledge, that disconfirms it. For example, studies using functional MRI have already evidence of detectable differences between the brain-states associated with the legally-relevant mental states of "knowledge" and "recklessness," respectively. *See* Iris Vilaris et al., *Predicting the*

a result, it would seem fair to conclude that there is a *three-way* correlation between (1) putatively causal mental states, such as intentions, (2) the co-occurring brain activity needed to produce the motor signals to the body, and (3) the bodily movements that occur.

Given such a three-way correlation among mental states, neuronal activity and movements, we are now left with the question of what causes what? In particular, are bodily movements caused by the mental states that are highly correlated with them or are they caused by the (also) highly correlated neuronal activity? The simple answer would seem to be that it is the neuronal activity alone that does the real causal work: It is the neuronal activity that causes the correlated bodily movement *and also* gives rise to or "enables"[195] the correlated mental states (which are, accordingly, dependent on it[196]). More elaborately, a commitment to the truth of physicalism and, in particular, of "causal closure"[197] would prevent choosing the mental correlate in preference to the physical (neuronal) correlate as the one that does the causal work—on the ground that every event that has a cause must have a sufficient physical cause. Then under an analysis sometimes known as the "exclusion principle," the mental cause must be "excluded" to avoid a double-counting of causes.[198] Therefore, if one wants to claim the mantle of physicalism, the question of causation within the three-way correlation has to be resolved in favor of neuronal causation; both the attendant mental states and the bodily movements would thus be dependent on the physical. There appears to be no way to

---

Knowledge/Recklessness Distinction in the Human Brain (2017), www.pnas.org/cgi/doi/10.1073/pnas.1619385114.

195. *See Common Sense*, *supra* note 46, at 47 ("[T]he brain enables the mind").

196. This idea that mental states are dependent on underlying neuronal activity for both their existence and their content is just another way of saying that the mind cannot directly apprehend and perceive facts about the outer world without the help of the sensory organs and their neuronal connections to the brain, and that it cannot float about freely above its physical base and autonomously generate trains of thought, intentions or other content on its own. If that is so (and a commitment to physicalism would seem to require it), then mental states can only have the content that is supplied to them by the biomechanical functioning of the brain, sense organs and nervous system. And if that is the case, it is difficult to see how the presence of particular mental states, such as intentions, could have any independent moral significance in attributions of responsibility. *See* PHYSICALISM DILEMMA, *supra* note 149.

197. *See* KIM, PHYSICALISM, *supra* note 15, at 36–45. "Causal closure" says that physical events must have sufficient physical causes (if they have any causes at all). The causal closure principle is further discussed in greater detail *supra* note 150.

198. According to the exclusion principle, it is inconsistent to say that an event has both a sufficient physical cause and also a necessary mental cause, and it is double counting to say that an event has both a sufficient physical cause and, *in addition,* a mental cause where the mental cause depends on the physical cause for its existence and content (viz. is supervenient on it). Events may, of course, have multiple sufficient causes in cases of genuine overdetermination, but it is not a case of genuine overdetermination when one of the multiple causes is dependent on the other. The exclusion principle is further discussed in greater detail in PHYSICALISM DILEMMA, *supra* note 149.

insist instead on the causal efficacy of mental states other than by either rejecting causal closure and the exclusion principle (which means disavowing physicalism) or by asserting that mental states can somehow float free of their underlying neuronal activity (also inconsistent with physicalism). Either of these moves would allow one to conclude that mental causation exists, but both of them would require a disavowal of physicalism.

In sum, there is no reason to doubt the "subjective experience of first person agency" that "virtually every neurologically intact human being takes for granted," but there is also no reason to interpret that experience as evidence of *mental* causation rather than purely neuronal causation. The foreknowledge one has that a bodily movement will occur (on which subjective sense of agency is neuronally based[199]) clearly supports an inference that the movement was caused by factors within oneself. But it would be pure conjecture to treat the sense of agency as amounting to a specification that those internal factors are mental as opposed to purely physiological.

Before closing, it should be stressed none of this is meant to say that the "experience" of mental causation is an illusion. The point is, rather, that there is no such thing as an "experience" of mental causation at all. What there is, instead, is an *inference* of mental causation that people draw based on a simple correlation between other experiences that they have (viz. their conscious intentions and their subsequent bodily movements). And is a fallacious inference at that. As a result, no matter how many times the intentions/actions correlation is observed and noted, it can no more show that mental states cause bodily movements than a stack of calendars can show that Tuesdays cause Wednesdays. Even less does the correlation provide a basis for asserting that mental causation is entitled to a *presumption* of truth. The neuroskeptic's case for mental causation is seriously overclaimed.[200]

B. The Detection Argument for Mental-State Causal Efficacy

In addition to the correlation evidence that is commonly (and fallaciously) cited as proof of mental causation, there is also what might be called a "detection" argument. The idea of the detection argument is this: It must be true that mental states can cause events to occur in the physical domain because, if they could not, there would be no plausible explanation for certain observed behavioral facts, such as the fact that people say they

199. *See supra* note 191.
200. *Cf. Overclaim*, *supra* note 46, at 397–412; *Neurolaw*, *supra* note 46, at 7, 31.

have mental states, talk about them and even write articles and books about them. There would be nothing to systematically cause the physical brain to produce these expressive behaviors (lips moving, fingers typing, etc.) if it could not detect the existence of mental states and be physically affected by them.

It is inferable from observed behavior not only that physical brains can detect the *existence* of mental states (i.e., consciousness) but also that they can detect the fact that mental states differ from one another in qualitative content, i.e., that they differ in what they are about[201] and in what it "is like" to have them.[202] This ability of the brain to detect that mental states are *diverse* can be reliably inferred from the observed behavioral fact that people describe their mental experiences as differing from one another, being apt to say, for example, that the experience of a pinprick does not feel the same as a spoonful of crème brulêe. If physical brains were not able to detect and be physically affected by the diverseness of mental states, the brain would have no cause to produce the observable expressive behavior of talking and writing about the differences. Accordingly, from the fact that physical brains do cause people to talk and write about both the existence of mental states and the diversity among them, one can conclude that both their existence and their diverseness are able to produce perturbations in the electrical patterns that process information in the brain.[203] At least to this extent, mental states appear to be causally efficacious in the physical domain.

The detection argument seems sound in its limited application (showing that brains can be affected by the existence and diverseness of mental states), but it would go beyond the evidence to say that the detection argument establishes that behavior can be caused or influenced by intentions and reasons, i.e., that persons are "agents." And as long as "no responsibility is possible if people are not agents,"[204] then the limited

---

201. For example, if you are consciously aware that you're seeing a duck, then you have a mental state *about* seeing a duck. If you're consciously aware that you are sipping pinot noir, you have a mental state *about* sipping pinot noir. If you are consciously aware that you intend to rob a supermarket, you have a mental state about intending to rob a supermarket. The "qualitative content" of such mental states include, as part of the overall mental state, a conscious awareness of the sight of a duck, of wine gliding across the tongue or of desiring to rob a supermarket, as the case may be. The idea of the qualitative content of mental states includes what are sometimes referred to as "qualia" and propositional attitudes.

202. *Cf.* Thomas Nagel, *What is it Like to Be a Bat? in* MORTAL QUESTIONS (1979).

203. Actually, it might also be possible that the behaviors concerning mental states described in this section might occur even if mental states did not exist at all as long as people have the illusion of mental states (whatever that might mean). I am setting that possibility aside, however, since people who think mental states are illusions surely do not believe in mental causation, and my aim is to show that the belief in mental causation, for those who have such a belief, is ungrounded.

204. *Inevitable Mind*, *supra* note 3, at 46; *Common Sense*, *supra* note 46, at 67; *see also supra*

mental causation established by the detection argument is not enough to justify blame and punishment.

The reason that the detection argument does not suffice to provide evidence of "agency" is this: In order for the brain to produce behavior reflecting intentions and reasons, the brain would have to be able to detect what those intentions and reasons are—meaning that mental states such as intentions and reasons would have to able to physically modify the brain or, at least, its patterns of electrical activity, *according to their qualitative content.* However, the only behavioral evidence that mental-state content (such as intentions and reasons) could physically modify the brain or the patterns of electrical activity in it is the introspectively observed fact that there is a correlation between intentions and reasons and subsequent actions. This correlation alone cannot, however, support a valid inference that the brain can detect the content of mental states because, in the absence of any suggestion of an explanatory mechanism, such an inference would be, once again, pure *post hoc ergo propter hoc.*

Besides, there is a perfectly good physical explanation for the observed correlations between mental-state content and subsequent actions, namely: Because mental states depend for their content on brain states,[205] any information that the brain could gain by detecting mental-state content would simply duplicate informational content that the brain already has. For example, if a motorist is prompted to stop when she approaches a red light, the pivotal causative trigger leading her to push the brake need not be the conscious mental experience of seeing redness. She could just as well be triggered to stop by the physical fact that the radiation entering her eyes from the traffic signal has a wavelength of roughly 680 nanometers rather than 520 nanometers (green). The neuronal detection of the radiation's wavelength is perfectly adequate to serve as the dispositive factor in the computational production of the resulting behavior, viz. brake-pushing rather than gas-goosing. There is no need to suppose that the brain would have need to refer to the content of any co-occurring mental states which, in any event, are epistemically dependent on the neuronal detection of the light's wavelength. In short, there is no reason why the brain cannot plausibly produce behavior that is highly correlated with the qualitative content of mental states using the very same information that the mental-states' content itself is based on—information that the physical brain collected and assembled in the first place.

To take another example, the more obvious reason why a person reacts differently to the soft fur of a kitten than to a hammer blow on the thumb

---

Part II.A and notes 98–99.

205. See *supra* note 196.

is not that the two events produce different mental-state content in consciousness, which is then detected by the brain. The more obvious reason is an ordinary physical reason, namely, that sensory data coming in from the peripheral touch receptors in the hand activate different populations of neurons in the case of the kitten than in the case of the hammer. The activation of those different populations of neurons gives rise to the different mental-state content that is experienced in the two cases as well as producing the different sets of behavioral reactions that follow.

One last point: If the brain manifestly can detect that mental states differ from one another in qualitative content (mental-state diversity), does it not follow that it must also be able to detect what that qualitative content is—what the various different mental states are like and what they are about? The answer is not necessarily. The brain could, for instance, be in much the same position as an archeologist who discovers a text written in an unknown ancient language: While the archeologist can easily detect that the text contains many different words, she cannot detect what the various words mean—their content. To be sure, it seems plausible to assume that the brain's computational machinery is able to *ascribe* qualitative content to different mental states by associating them in memory with the various sensory perceptions, proprioceptions, and states of neuronal activity that co-occurred with them. And such ascribed contents of remembered mental states should align closely with the mental states' actual contents.[206] It is even plausible that memories of content previously ascribed to mental states experienced in the past could become factors in the brain's computational decisions when producing behavior later on. But even if all this does indeed occur, it still would not be a true case of mental-state content causing effects in the physical domain. It would still be the brain's physical processes alone that are doing all the causal work.[207]

---

206. Still assuming, as before, that mental states depend on brain states for their content. *See supra* note 196.

207. The argument in this (and the preceding two) paragraphs owes a great debt to Jaegwon Kim, who has presented similar ideas in somewhat different terms. *See* KIM, PHYSICALISM, *supra* note 15, at 170–73. According to Professor Kim, if I understand him correctly, the brain can easily detect that red lights and green lights are different even if it has no information at all as to what it is "like" to see red or green. The reason is that different light wavelengths cause different raw data to be sent from the eyes. While it often occurs that behavior is modulated by detected *differences* in what it is "like" to experience differing non-reducible mental events, such as color differences, it does not appear that there are instances in which behavior is modulated by the intrinsic nature of what it is "like" to see a given color (or to experience other "qualia"). *See generally id. But cf.* Mørch, *supra* note 9 (for a strong argument using as counter-examples the qualia of pain and pleasure).

In sum, while the detection argument for mental causation may show that the existence and diverseness of mental states have effects in the physical domain, the evidence does not support a valid inference that the qualitative content of mental states (intentions, reasons, etc.) can alter the patterns of electrical activity that process information in the brain. It does not, in other words, support a conclusion that such mental-state content can ever *in se* make a difference in the physical domain—as is said to be required as a prerequisite for agency and responsibility.[208] To infer that the content of mental states can make a physical difference is to make an inference based on bare correlation, an inference that would be premised on the logical fallacy of *post hoc ergo propter hoc*.

V. INFERENCE TO THE BEST EXPLANATION

Thanks to the findings of neuroscience in recent decades, we now have two different causal explanations of human behavior: First, there is the traditional, folk psychology view that the things we do are caused, at least in part, by mental states such as intentions and reasons. Second, there is the neuroscience alternative, viz. that human behavior is determined solely by ordinary physical causes acting upon the biomechanical physiology of the brain. The two causal explanations have very different implications for justice. The older, folk psychology explanation supports the widely shared position that persons who do wrong "deserve" to suffer because they are agents and therefore are morally responsible for their actions. The neuroscience alternative does not support this idea of "just deserts." Assuming that criminal justice practices are defensible only if their underlying factual premises are true, the crucial question is: Which of these two rival explanations of human behavior comes nearest to the truth? They cannot both be correct. How do we decide between them?

Abductive reasoning, or "inference to the best explanation," is a reasoning methodology that is commonly used to decide among competing inferences that can be drawn from a given set of observations and data.[209] The idea is to determine which of the competing inferences best explains the relevant evidence by comparing each one "with its rivals in point of explanatory power."[210] The inference that best explains the totality of the

208. *See Inevitable Mind*, *supra* note 3, at 46; *Common Sense*, *supra* note 46, at 67; *supra* Part II.A and notes 98–99.

209. *See* Gilbert Harmon, *The Inference to the Best Explanation*, 74 PHIL REV. 88–99 (1965). *See also* KIM, PHYSICALISM, *supra* note 15, at 126–30. For an excellent summary of the abductive reasoning process as applied to fact-finding in law, see Ronald J. Allen & Michael S. Pardo, *Relative Plausibility and Its Critics* 12–34 (2018), http://ssrn.com/abstract=3179601. *See generally* Igor Douven, *Abduction*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2017).

210. KIM, PHYSICALISM, *supra* note 15, at 129.

relevant evidence and data is inferred to be the one that comes closest to the truth.[211] It is a reasoning methodology that people use continually and intuitively, not just in science and law but in everyday life, to make sense of incomplete information or data. To take a simple example, suppose someone looks out her window and sees passersby carrying rolled up umbrellas. Two possible inferences occur to her. The first is that the day's weather forecast includes a prediction of rain. The second is that it is "Take Your Umbrella to Work Day." While either of these inferences could logically explain the observations and data at hand, she reflects that, though it often rains, she cannot recall hearing of a "Take Your Umbrella to Work Day." Besides, it is cloudy. She quickly decides which of the possible inferences provides the best explanation and prepares herself for rain.

It is not hard to see why, before the advent of modern neuroscience research, the inference of mental causation was, despite its frothy factual basis, accepted as the best explanation for human behavior. After all, what else was there?[212] During the long pre-history of neural research, when even the best informed scientists had no idea how brains worked or neuronal activity could produce human behavior, the mental causation hypothesis must have seemed obvious—as indeed, to many people, it still does today.[213] More to the point, when lawyers and judges were working to structure the criminal law's foundational assumptions several centuries

---

211. "[O]ne infers, from the premise that a given hypothesis would provide a "better" explanation for the evidence than would any other hypotheses, to the conclusion that the given hypothesis is true." Harmon, *supra* note 209, at 89.

Needless to say, the explanation that is to be considered "best" is the one that best accounts for the available evidence and data and not, for example, the one that best supports a particular policy preference. Also, it should be specified that, for present purposes, we are looking for the best *causal* explanation and, as noted earlier, not all explanations are "causal" explanations. *See supra* note 15. For some purposes, a non-causal explanation may be good enough or, even, preferable--for example, one might explain to a child that birds build nests because they need a place to raise their young rather than explain the actual causes (season-induced hormones, etc.). *See id*. Similarly, explaining our own and others' behavior in terms of mental states may, as Quine said, be useful as "irreducibly mental ways of grouping physical states and events." QUINE, *supra* note 15, at 72. Indeed, the common practice of explaining persons' actions by reference to accompanying mental states would be incontestably convenient whether or not mental states have a causal role. The reason is that everyone learns to "read" others' presumed mental states by associating them with behavioral clues, projecting from what they know of their own mental states. Having learned these associations, it then seems natural to explain and understand our own and others' behavior by reference to mental-state surrogates or metonyms instead of naming the underlying brain states. Such metonymic or surrogate explanations do not, however, mean that the mental states *actually* cause the behavior, at least not in the sense that would be required if mental causation is deemed to be an indispensable pre-condition for holding people morally responsible and deserving of punishment. It is not, in other words, the best *causal* explanation.

212. There were, to be sure, explanations that were even more far-fetched. *See supra* note 129.

213. *See supra* note 173.

ago, mental causation was not only the best available explanation of human behavior, it was the only game in town. As such, it prevailed by default.

> [However,] part of what lends credibility to the talk of inference to the best explanation . . . is the fact that putting competing theories to test on the basis of how well they explain the data is an ongoing, in-principle never-ending, affair. When further data are in, the rankings of the theories in terms of their explanatory power may very well change.[214]

In other words, one cannot simply say that, just because mental causation was the best explanation for human behavior 300 years ago, it still remains the best explanation today. It cannot be said that the question has been settled once and for all.

So how do the two alternative accounts of human behavior compare today in their ability to explain the totality of the relevant evidence and data? The neuroscience view, that behavior is the result of mundane physical forces channeled in ordinary physiological processes, is (unsurprisingly) consistent with and supported by the wealth of physiological evidence and data that decades of neuroscience research have produced. Based on that evidence and data, neuroscientists are able to provide sophisticated descriptions of the electrochemical activity in the brain and how it can cause the bodily movements that add up to human behavior. In literally millions of experiments over the past several decades,[215] neuroscientists have accumulated a deep, complex and internally consistent web of evidence and data showing that bodily movements, not just of people but of every motile creature on earth, have ordinary physical causes: In every known motile creature, neuronally-embodied information from the sense organs travels to the brain where it is processed along with previously registered information (memories) to create representations of potential behavioral options which are then computationally compared, leading to the generation of motor signals to the muscles to produce behavior that is calibrated to respond to the current outside situation in light of the individual's internal needs and memories of what worked in the past.[216]

In contrast to the broad and robust factual basis for the neuronal-causation thesis, the mental-causation hypothesis provides no description

---

214. KIM, PHYSICALISM, *supra* note 15, at 130.

215. *See supra* note 4.

216. *See* PASSINGHAM, *supra* note 4, at 67–70 (discussing how options are comparatively valued in the prefrontal cortex based on inputs from the various sensory perception area (for sight, sound, touch, etc.). *See generally* SAPOLSKY, *supra* note 6, at 21–77.

whatever of how incorporeal thoughts, flashing through the mind, could cause physical bodies to move. And this is not to mention the difficulty of fitting the mental-causation hypothesis in with the mass of hard evidence produced by neuroscience.

Perhaps in recognition of this latter difficulty, there is a tendency among neuroskeptics either to ignore the implications of the neuroscience (especially in legal discussions) or to pre-emptively dismiss the science on the ground that it fails to answer certain non-germane questions, such as how neuronal activity brings about mental states or how mental states produce actions.[217] Professor Morse writes, for example, that it is "neuroarrogance" to suggest we should consider making big changes in punishment practices based on neuroscience "[g]iven how little we know about the *brain-mind* and *brain-mind-action* connections."[218] Professor Morse is, of course, quite correct in saying that today's neuroscience does not does not explain the connection between the brain, the mind and actions—the "*brain-mind-action*" connection. His mistake is to think it matters.[219]

As far as neuroscience's causal explanation of *behavior* is concerned, the brain-mind-action connection that Professor Morse finds so vital is a total irrelevancy. The neuroscience explanation is entirely physiological, and the mind as such has no evident role to be explained. On the contrary, for neuroscience the only real question about behavior is the "brain-body" question: how the brain produces movement. And in answer to *that* question, neuroscience has data in abundance. It supports an explanation that is based entirely on ordinary physical forces and well-understood biochemical and electrical causes. It is an explanation that has no place for mental causation as part of the explanatory structure.

---

217. Morse, *LawSci*, *supra* note 37, at 23–24; Morse, *Common Sense*, *supra* note 46, at 67; Morse, *Exuberance*, *supra* note 46, at 859. Dennis Patterson, *Neuromania: A Review of Peter A. Alces, The Moral Conflict of Law and Neuroscience* 5 J. L. BIOSCI. 3 (2018) (book review) ("[N]euroscience simply has not progressed to the point where it can even tell us how the brain enables the mind"); The findings of neuroscience are also dismissed on the ground that they are irrelevant to normative questions because they can only show causes and "causes are not excuses." *See supra* note Part II.D.

218. *See* Morse, *Overclaim*, *supra* note 46, at 397–412 (emphasis added); *see also* Morse, *Neurolaw*, *supra* note 46, at 7, 31; Morse, *Overclaim Redux*, *supra* note 46.

219. Notably, perhaps, it does *not* seem to matter to the mind-centric friends of mental-causation that their alternative likewise provides no explanation the *brain-mind* and *brain-mind-action* connections, much less a "better" explanation. As a criterion of "arrogance," the ability to account for these connections seems to receive decidedly selective application.

As long as there is no evidence that mental states and consciousness play a causal role in determining what people do (and there is not), neuroscience has no need to explain them or take them into account in explaining the causes of human behavior. To be sure, the creation and ontological nature of conscious mental states may be of enormous philosophical interest, but as a causal explanation of *behavior*, the neuroscience explanation is not diminished in the slightest by the fact that mental states remain "conceptual and scientific mysteries."[220]

This is not to say, of course, that the neuroscience explanation of behavior has been fully fleshed out or there is no more to know about it.[221] But despite the existence of many areas for further research, the neuroscience evidence we already have leaves no genuine question that mechanical, computational processes alone are capable of generating bodily movements—as they do throughout the animal kingdom. From studies of animals having brains far less evolved than ours (and, almost certainly, no minds), we know with veritable certainty that sophisticated behavioral repertoires can be managed by neuronal activity alone, with no plausible possibility of control by mental states. Even insects are known to rescue injured comrades and nurse them back to health.[222] And the world is packed full of mindless creatures that are fully able to set themselves into motion and do all they need do to survive and thrive under the direction of mechanical computational processes that are dynamically responsive to their situations and neuronally-embodied behavioral dispositions. Even though there are still some gaps in the neuroscience story, there are none that "mental causation" (about which we know nothing at all) could helpfully fill.

To summarize, in a world of less-than-perfect information, the issue is not whether an explanation is perfectly complete or "conclusive." The issue is, rather, which of the inferences from the facts we know provides the best explanation for the totality of relevant evidence and data.[223] On this score there is really no contest between mental-causation explanation of behavior and the neuroscience explanation. The biomechanical description provided by neuroscience is both robust and supported in

220. Morse, *Inevitable Mind*, *supra* note 3, at 33.

221. *See* Morse, *LawSci*, *supra* note 37, at 24 ("[W]e still do not have sophisticated causal knowledge of how the brain enables the mind and action generally"). *See also supra* note 34.

222. Erik T. Frank et al., *Wound Treatment And Selective Help In A Termite-Hunting Ant*, PROCEEDINGS OF THE ROYAL SOCIETY B. (Feb. 14, 2018), http://rspb.royalsocietypublishing.org/content/285/1872/20172457.

223. *See* KIM, PHYSICALISM, *supra* note 15, at 129 ("[If the] aim is to reach a conclusion that we should believe as true, or be prepared to use as a guide for action, [we] must respect the principle of total evidence, and in the present case this means that the data, or evidence, to be explained must be all the data relevant to the issue at hand").

detail by an intricate and internally coherent web of experimental evidence and data. The mental-causation hypothesis, by contrast, merely posits a blunt causal relationship between two widely disparate kinds of events, mental and physical, suggesting nothing in the way of a mechanism by which that causal relationship could be effectuated.[224]

Even more importantly, the mental-causation hypothesis does not deal with or even try to integrate the parallel track of behavior causation that neuroscience has revealed. Inferring mental causation from the fact that actions follow intentions is not merely fallacious in itself but it provides no explanation whatever of the extant evidence and data. It neither explains them away as false (a dubious possibility) nor does it suggest how mental states could override the physical brain's behavioral determinations and thus secure a distinctive causal role that is consistent with the evidence and data.

If it is the "best" explanation that we are looking for, the one that best fits together all of the relevant evidence and data, the mental-causation hypothesis, rooted in logical fallacy, is no longer the one. The best explanation is that neuronal activity, not intentions, reasons and other mental states, are the actual causes of what people do.[225]

VI. SOME IMPLICATIONS FOR JUSTICE

The accumulating wealth of findings from neuroscience research makes it increasingly clear that current criminal justice practices are morally predicated on a factual premise that is almost certainly false, namely, that wrongful conduct is caused by intentions, reasons and other mental states. From this false premise has followed the faulty moral certitude that those who do wrong "deserve" to endure suffering and hardship at the hands of the state. In view of the evidence we now have,

224. In demanding a complete explanation and "conclusive" proof, *see supra* text accompanying notes 105–112, the neuroskeptic who rejects the "best explanation" of human behavior is like the person who does not "completely" understand cars and therefore feels entitled to reject ordinary physical explanations. Though he knows about pistons, spark plugs, gasoline and so on, there is a lot he does not know, so he says: "Until someone shows me conclusive proof of the 'motor' explanation of automobile movements, I'm entitled to believe that the car moves because I hold my hands on the steering wheel and my foot on the pedal (*post hoc ergo propter hoc*)." Indeed, in the absence of a "conclusive" explanation, I suppose that one *is* "entitled" chose any explanation one pleases—but that is not the question. The question is whether one is entitled to use that explanation as a part of a pretext for inflicting hardship and deprivation on millions of others.

225. Note that to say that mental states like intentions and reasons are not effective causes of human behavior does not by any means equate to saying they do not exist at all. I see no basis in the evidence of neuroscience for saying that they do not. I am quite sure that they do exist, but (like most things) they may not be what we think they are. The "thing in itself" of intentions, reasons and reasoning does not necessarily mirror the *phenomena* of them in consciousness.

however, the "best" explanation of human behavior—the one that best accounts for the totality of the evidence—is the brain-based biomechanical explanation that is provided by neuroscience. The older but still dominant mental-causation hypothesis, premised on logical fallacy, is no longer credible (for those who demand evidence) and in any case is no longer necessary to a complete factual explanation of why people do what they do. Based on today's best explanation of human behavior, there is good reason to believe that current punishment practices and our bloated criminal justice system[226] can no longer be morally justified.

The treatment of offenders in our criminal justice system systematically deprives millions of Americans of practically everything makes life worth living, and it is sharply biased racially to boot. It is designed to be callous and indifferent to the distress and misery of the people subjected to it, and it is. Given the immense human cost of this system, not just to those actually incarcerated but also to their families and their communities, the intellectual bankruptcy of its justificatory underpinnings should be a matter of serious concern.[227] To the extent that the justification for current punishment practices depends on the belief that offenders "deserve" what they get because of their own bad intentions, the cloud cast by neuroscience on the mental causation hypothesis unavoidably prompts the question: What are the implications for justice of recent neuroscience findings that show it is extremely unlikely, based on the totality of evidence, that wrongful conduct is in fact caused by intentions, reasons or other mental states?

One major possibility is, of course, that recent neuroscience findings will turn out to have *no* impact on criminal justice practices or the treatment of offenders. An appetite for retribution runs strong in human psychology, and people become indignant, even angry, when someone suggests that offenders should not be punished for their crimes.[228] The

---

226. The United States has the highest prisoner population in the world and the highest number per capita (with stark racial disparities in rates of incarceration). *See Countries With the Largest Number of Prisoners Per 100,000 of the National Population, as of July 2017*, STATISTA, https://www.statista.com/statistics/262962/countries-with-the-most-prisoners-per-100-000-inhabitants/; Institute for Criminal Policy Research, *World Prison Brief*, http://www.prisonstudies.org/highest-to-lowest/prison_population_rate; John A. Humbach, *Is America Becoming a Nation of Ex-Cons?* 12 OHIO ST. J. CRIM. L. 605 (2015).

227. *See generally* Andrew M. Koppelman, *American Evil: A Response to Kleinfeld on Punishment*, 50 ARIZ. ST. L.J. 179 (2018); Christopher Wildeman et al., *Conditions of Confinement in American Prisons and Jails*, ANNUAL REV. L. & SOC. SCI. (2018), https://doi.org/10.1146/annurev-lawsocsci-101317-031025).

228. *See* Isaac Wiegman*, The Evolution of Retribution: Intuitions Undermined,* 98 PACIFIC PHIL Q. 193 (2017)*;* Katherine Harmon, *Does Revenge Serve an Evolutionary Purpose?* SCIENTIFIC AM. (May 4, 2011) ("Magnetic resonance imaging (MRI) scans have revealed that thinking about revenge activates the reward center—where the feel-good neurotransmitter dopamine is lodged—in much the same way that sweet foods or even drugs can"), https://www.scientificamerican.com/articl

suffering of others can be a soothing consolation for suffering of one's own.[229] Even when science provides facts that show we are better off if we sublimate vestigial yearnings,[230] people's understanding of their world "is not just a matter of arguments" but what they "want to be true."[231] The findings of science can simply be ignored. "Human kind," wrote T.S. Eliot, "cannot bear very much reality."[232]

Another possibility is that the findings will be accepted but the law could be adapted in order to continue the old ways. One conceptually easy adaptation would be simply to revert to the idea that requital alone is a sufficient basis for deserving punishment.[233] Alternatively, it would be a simple matter to modify the legal elements of "intention" and "volition" so they do not refer to mental states but instead to brain states—specifically, to those brain states that are evidenced by behavior manifestations that the law already uses to infer intentions and other mental states.[234] The Supreme Court has shown itself perfectly willing to approve replacing a legal mens rea requirement with a behavioral surrogate as long as the defendant manifests behavior that mimics the prescribed mens rea.[235] In any case, it is the moral logic of democracy that, in general, the people's desires must be accorded great deference because there is no earthly

---

e/revenge-evolution/. See also my discussion of this topic in John A., Humbach, *The Humane Principle and the Biology of Blame (Evolutionary Origins of the Imperative to Inflict)*, *in* PROCEEDINGS OF 3RD ANNUAL GLOBAL CONFERENCE ON PERSPECTIVES ON EVIL AND HUMAN WICKEDNESS (2003), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1524257.

229. Or maybe it is just that "we love this kind of thing" and have a great yearning to hurt someone, and they pounce on the opportunity when it arises." Koppelman, *supra* note 301, at 187–88.

230. As, for example, nutritional science supports efforts to resist the now counter-adapative predilection for dietary sugars and fats.

231. Larissa MacFarquhar, *Mind Expander*, THE NEW YORKER (Apr. 2, 2018) (describing thought of Scotland-based philosopher Andy Clark) ("[T]he way you understand yourself and your relation to the world is not just a matter of arguments: your life's experiences construct what you expect and want to be true"); *see also* Jane Esberg and Jonathan Mummolo, *Explaining Misperceptions of Crime,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3208303 (exploring reasons for the disconnect between public beliefs about crime and the actual facts of the matter concerning the prevalence of criminality).

232. T.S. Eliot, *Burnt Norton*, *in* FOUR QUARTETS (1943).

233. As noted earlier, studies have shown that, as an abstract matter, people say they think determinism is not consistent with holding offenders responsible but, when presented with concrete cases of bad behavior, they tend to attribute responsibility even for actions caused by neurological disorders. *See* Nichols, *supra* note 77, at 1402.

234. Already, intentions and the like are inferred from behavior, since mental states cannot be known directly. *See, e.g.,* People v. Conley, 543 N.E.2d 138 (Ill. App. 1989). It has been argued that this is close to what courts have always implicitly done anyway. *See* Katrina L. Sifferd, *supra* note 9, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512325. ("When a judge or jury is looking for the mental state that is causally linked to criminal harm, they are also implicitly looking for whatever physical state realizes the content of that mental state.").

235. Egelhoff v. Montana, 518 U.S.37 (1996) (premeditation requirement satisfied if a defendant too drunk to premeditate *acted* like he premeditated).

authority with legitimate power to overrule them. If the democratic "will" demands placatory torture[236] in the name of justice, the demand may be its own justification.

Let us assume, however, that the neuroscience facts turn out to matter, at least in the longer run, and that an effort is made to align criminal justice policy with moral justifications that are predicated on the best explanations of the evidence and data. Let us assume that, when the mental causation hypothesis is factually discredited, the idea that people "deserve" to suffer will lose favor and *neminem laede*[237] will replace retribution and government-endorsed hate[238] as our fundamental principle of justice.[239] How might criminal justice practices be different in such a "brain-compatible system [that] prizes fairness and long-term crime prevention over harsh but inconsequential punishment"?[240]

Obviously, if the retributive rationale for punishing presupposes mental causation as the basis for blame, then dissolving the mental-causation creed would leave retribution as "the mere addition of a second evil" to the one that has already happened.[241] It would no longer make moral sense. Unless some other basis could be found for concluding that offenders "deserve" to suffer (based, perhaps, on faulty empirical character[242] or low "virtue"[243]), continued inflictions of hardship and deprivation in the name of retribution would be bereft of their factual underpinnings and have to be seen as flatly immoral.

The utilitarian rationales for punishment (e.g., to deter, to incapacitate or to rehabilitate) would probably not, as a practical matter, fare much better. It is true that discrediting mental causation would not affect the applicability of the basic utilitarian rationale for punishment, namely, that inflictions are morally justified as long as they produce benefits to others that outweigh the detriments to those afflicted: The ends, as it were, justify

---

236. Making people suffer in order to appease others who want the suffering to be inflicted.

237. Injure no one.

238. For a rare honest acknowledgement, *see* 2 JAMES FITZJAMES STEPHEN, A HISTORY OF THE CRIMINAL LAW OF ENGLAND 81 (Cambridge Univ. Press 2014) (1883) ("The criminal law thus proceeds upon the principle that it is morally right to hate criminals, and it confirms and justifies that sentiment by inflicting up on criminals punishments which express it").

239. ARTHUR SCHOPENHAUER, ON THE BASIS OF MORALITY 149 (Eric. F.J. Payne, trans. 1995) [hereinafter SCHOPENHAUER, ON THE BASIS]; *see also* SCHOPENHAUER, THE WORLD, *supra* note 171, at 343 ("[T]o diminish the suffering spread over all . . . the best and only means [is] to spare all men the pain of suffering wrong by all men's renouncing the pleasure to be obtained from doing wrong").

240. David M Eagleman and Sarah Isgur Flores, *Defining a Neurocompatibility Index for Criminal Justice Systems: A Framework To Align Social Policy With Modern Brain Science*, *in* LAW OF THE FUTURE 161–72 (2012).

241. SCHOPENHAUER, THE WORLD, *supra* note 171, at 350.

242. *Cf.* ARTHUR SCHOPENHAUER, PRIZE ESSAY ON THE FREEDOM OF THE WILL 86 (Eric. F.J. Payne, trans. 1999).

243. I have written about this general possibility in a working paper entitled *Free Will Ideology* (2010), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1578445.

the means. But even though utilitarianism's goal-driven justifications for punishment can be served irrespective of blame or just deserts, there is reason to think that most punishment-utilitarians in real life tacitly assume that it is unjust to punish the innocent.

There is, after all, something prima facie dubious about the idea that it is morally acceptable for people "to use another's suffering as a means to the attainment of [their] ends."[244] Few would agree, I think, that it is fine to torment Alice just because Ray and his friends will benefit to an even greater degree. More to the point, it is seldom heard that governments should punish the blameless in order to prevent harms to other innocents, for example, by afflicting the families of offenders with a view to deterring the offenders themselves.[245] And if benefit to others could justify punishment even in the absence of desert, a committed utilitarian should not care whether criminal trials are fair in their determinations of guilt, or even try to be. All that should matter to the committed utilitarian is that trials do a good job of estimating how much net benefit others will receive if the government subjects the defendant to punishment. It is, however, a rare utilitarian who argues that it is pointless to make criminal trials fair and that the law should concentrate instead on maximizing overall utility. In short, even if punishment is meant for the greater good of others, few are so committed to utilitarian ideals that they would inflict it on those who do not somehow "deserve" it. And to the extent that real-life utilitarians tacitly fall back on the notion that punishment must be reserved for the guilty, the cloud that neuroscience casts on mental causation should seriously compromise the utilitarian case for punishing anyone at all.[246]

Still, there are other facts that cannot be ignored. First and foremost, there is the fact that every society contains individuals whose behavioral

---

244. SCHOPENHAUER, ON THE BASIS, *supra* note 239, at 149.

245. Such derivative or vicarious punishments are certainly not unheard of. *See e.g.*, NUM 14:18 (New International Version) (The Lord "does not leave the guilty unpunished; he punishes the children for the sin of the parents to the third and fourth generation"); *Prisons of North Korea - Camp 14 Kaechon*, U.S. DEPT. OF STATE BUREAU OF DEMOCRACY HUMAN RIGHTS AND LABOR (2017), https://www.state.gov/j/drl/rls/fs/2017/273647.htm. And, indeed, imposing punishment on offenders' families may be the only effective way to deter certain crimes, such as suicide attacks or defections to the enemy. *See* Anna Ahronheim, *IDF Destroys Homes Of Four*, JERUSALEM POST (Aug. 10, 2017), http://www.jpost.com/Arab-Israeli-Conflict/IDF-destroys-homes-of-four-Palestinians-responsible-for-deadly-terror-attacks-502062.

However, the United States Constitution arguably disallows such third-party punishment under the rule that forbids punishing any person who neither pleads guilty nor is proved guilty of a crime beyond a reasonable. *See In re* Winship, 397 U.S. 358 (1970).

246. In any case, I seriously doubt whether, when it comes to punishment at least, most ordinary people are actually utilitarians anyway. That is to say, I do not think most people really care too much whether the benefits of punishment exceed the costs, or vice versa, as long as the costs are mostly borne by somebody other than themselves (such as offenders and their families). They just want to be protected from crime.

dispositions pose unreasonable risks of harm and danger to others.[247] Moreover, given that most people have powers of self-control, in varying degrees, lawmakers would be seriously remiss if they did not design the law to make the most of these powers with a view to minimizing harms.[248] Nothing contained in the neuroscience explanation of human behavior denies these truths. Nor does anything in neuroscience deny that coercive intervention is sometimes necessary to prevent or minimize risks and harms[249]—not merely as a conventional moral imperative but as a central part of the state's raison d'etre.[250] I would certainly want and expect the state to intervene coercively to protect me and my family from hardship or deprivation at the hands of dangerous others who would inflict it, and I am confident that most would agree. But it does not follow that the people who must be coerced by the state in the interest of public safety morally "deserve" such treatment. Even though suffering and adversity are inseparable from reducing behavioral risks of harm to reasonable levels, it does not follow that any person should suffer more than necessary to achieve this goal.[251]

If retribution and utilitarian goals do not justify the suffering that is inseparable from coercive force for public protection, what does? How can one justify such suffering cast on risky or dangerous individuals if it is not deserved? While it is not the ambition of this article to lay down a definitive moral foundation for the coercive treatment of dangerous individuals or suggest new models of societal protection to replace traditional "punishment," I also do not want to leave the impression that no possibilities exist. Accordingly, let me offer the following example of a justification for coercive treatment that is not based on a presupposition of just deserts.

---

247. That is to say, individuals whose brains (patterns of synaptic strengths) have been modified by past experiences in interaction with their genetic and developmental endowments, in such a way that they have become exceptionally disposed to respond to certain combinations of external circumstances and other extra-cerebral factors in ways that pose unreasonably high risks of danger to the interests of others. (For this purpose, "unreasonably high" is understood to mean so unusually high as to be too socially intolerable to allow—determined as a matter of public policy in accordance with law.)

248. *See* Patricia Churchland, *The Big Questions: Do We Have Free Will?*, NEW SCIENTIST 42 (Nov. 2006).

249. *See, e.g.*, Robert M. Sapolsky, *The Frontal Cortex In The Criminal Justice System,* 359 PHIL. TRANS. ROYAL SOCIETY BIOLOGICAL SCIENCES 1787 (2004) ("[Y]ou do not ponder whether to forgive a car that, because of problems with its brakes, has injured someone, [but] you nevertheless protect society from it").

250. *See* U.S. CONST. pmbl. ("[I]n order to . . . establish justice, insure domestic tranquility").

251. The more fundamental principle is equality—that that every person is equally morally deserving of decent treatment and respect, there being no normative basis (absent mental causation) to hold any person more (or less) morally deserving than any other.

My example of a morally defensible theory for justification of coercive force starts from the fundamental principle closely resembling that which underlies private self-defense,[252] namely: "[W]hat is done simply in order not to suffer any wrong is not wrongdoing."[253] To elaborate at bit, since it is wrong to harm or threaten harm to others by violence or cunning, no one has a right to do so. Therefore, to "resist" such encroachments[254] by "warding [them] off cannot itself be wrong" even though "violent action committed in connection with it."[255] Warding off is "justified . . . by its motive."[256] To wrongfully encroach on the "sphere" of another by violence or cunning is a denial of the other's equality, so warding off that encroachment is only the denial of that denial.[257] And one has "a *right* to deny that other person's denial with what force is necessary to suppress it."[258] The force used in warding off or resisting another may permissibly exceed the force used in the original encroachment,[259] but there is a built-in limit: If the force used against an encroacher exceeds that which is necessary for warding off, the excess would it would be a wrongful encroachment on the original encroacher. Uses of force that "transgress this limit, [are] consequently wrong, and [they] can therefore in turn be warded off without wrong."[260]

---

252. *See* Gregg D. Caruso, *Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model*, 32 S.W. PHIL REV. 25, 28–31 (2016); Derk Pereboom, *Free Will Skepticism and Criminal Punishment*, *in* THE FUTURE OF PUNISHMENT 49 (Thomas Nadelhoffer, ed. 2013). While I think the analogy of self-defense is very illuminating, my reason for seeking the principle underlying self-defense rather than invoking self-defense itself is that self-defense as practiced is not a "moral primitive" (an irreducible principle needing no further explanation) but a constructed contingent right that depends on factors, such as imminency and freedom from fault, that are probably germane to the protection of public safety. *See* United States v. Peterson, 483 F.2d 1222 (D.C. Cir. 1973).

253. SCHOPENHAUER, THE WORLD, *supra* note 171, at 342; *see generally id.* at §62, especially 339–42. In the passage that follows, my borrowing from Schopenhauer, who wrote over 150 years ago, is selective and it does not mean that I necessarily buy into a number of his related views concerning punishment, such as his contractarianism, e.g., *id.* at 347–349, or his view that the "object of punishment . . . is deterrence." While I think that the latter reflects a well-motivated rejection of retribution (calling it "wickedness and cruelty [that] cannot be ethically justified", *id.* at 348), it seems to me that *general* deterrence is also hard to justify if offenders cannot be shown to "deserve" suffering, and *special* deterrence would go beyond the "need to ward off" described in the text unless it is limited to mild *disincentivizing* or, perhaps, as a last resort when less encroaching measures would be ineffective.

254. SCHOPENHAUER, ON THE BASIS, *supra* note 239, at 154.

255. *Id*.; SCHOPENHAUER, THE WORLD, *supra* note 171, at 339.

256. SCHOPENHAUER, THE WORLD, *supra* note 171, at 339.

257. *Id.* at 339–40. I have avoided Schopenhauer's references to the metaphysical "will" because inclusion would likely be confusing for those unfamiliar with his overall system. I believe, however, the resulting paraphrase is faithful to his meaning.

258. *Id.* at 340 (emphasis in original). Schopenhauer specifies that cunning may also be used to oppose another's violence—so-called "white lies."

259. *Id.*

260. *Id.* at 340.

That's the principle: that it is not wrong to ward off wrong. Obviously, it is not meant as an answer all moral questions[261] but is suggested here as an example of a moral kernel for a theory of coercion can take the place of retribution and deterrence as a justification for public coercion—one that does not depend, openly or tacitly, on putative "just deserts." And that is its virtue. Even though governments sometimes still would have to treat people coercively, abandoning the thesis that some people actually "deserve" hard treatment, because of their mental states, should make an enormous difference in punishment practices.

To be sure, imprisonment facilities would surely still need to exist, for incapacitation and its incidental deterrent effects as well as places for rehabilitation.[262] But the mission of such facilities and, indeed, of the whole criminal justice system would be crime-prevention, not punishment. Actual incarceration would be treated as a last resort not the default response[263] and, most importantly, facilities would be designed to minimize rather than to accentuate the inevitable hardship and deprivation

---

261. Of particular importance, it elides the question of how the putative equality of a state of nature can be affected as duly constituted governments make laws, such as laws authorizing acquisition and control of tangible things and intangible values (property), which can radically change scopes of protected "spheres" and, hence, the line between right and wrong. This is not the place for such a discussion because questions of appropriate law enforcement and punishment *presuppose* the existence of norms to be enforced. I will only note that, in general, laws dealing with protection of the person support and further equality while those that protect property generally have the opposite effect, creating and maintaining *in*equality—though for purposes that are arguably sufficient to make them just. Laws that prescribe the disposition of offenders are, however, exceptions to this distinction: Although they are laws that deal with protection of the person, they generally operate by reducing the protection afforded to other persons (VNC transgressors) and, thus their immediate effect is to increase *in*equality.

262. There is evidence that imprisonment reduces recidivism, but only if it "increases participation in programs directed at improving employability and reducing recidivism," whereas for those who already have positive employment histories, imprisonment tends to *increase* recidivism. *See* Manudeep Bhuller et al., *Incarceration, Recidivism, and Employment,* NHH Dept. of Economics Discussion Paper No. 14/2018 (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id =3205006. In the United States, however, prison generally *decreases* employability. See John Schmitt & Kris Warner, Center for Economic and Policy Research, *Ex-Offenders and the Labor Market*, 12-13 (2010), http://www.cepr.net/documents/publications/ex-offenders-2010-11.pdf; Steven Raphael, *The Employment Prospects of Ex-Offenders*, 25 FOCUS 21 (2007), http://www.irp.wisc.edu/publications/ focus/pdfs/foc252d.pdf; Bruce Western, *The Impact of Incarceration on Wage Mobility and Inequality* 67 AM. SOC. REV. 526 (2002), http://scholar.harvard.edu/files/brucewestern/files/western_asr.pdf.

No position is taken here on the justifiability of rehabilitative interventions by the state, especially neuro-interventions, designed to change aspects of a person's personality against her will. *See* Christoph Bublitz*, "The Soul is the Prison of the Body"—Mandatory Moral Enhancement, Punishment & Rights against Neuro-Rehabilitation*, *in* TREATMENT FOR CRIME: PHILOSOPHICAL ESSAYS ON NEUROINTERVENTIONS IN CRIMINAL JUSTICE (David Birks & Thomas Douglas, eds., 2017). I would, however, agree there is no moral right to pose an unreasonable risk of harm to others.

263. Judges in the Netherlands, by using "alternatives to prison such as community service orders, fines and electronic tagging of offenders," have contributed to a forty-three percent reduction in the number of people incarcerated over a period of about ten years. *See* Lucy Ash, *The Dutch Prison Crisis: A Shortage of Prisoners*, BBC NEWS (Nov. 10, 2016), http://www.bbc.com/news/magazine-37904263.

that comes with loss of freedom for the protection of others.[264] A high priority would be placed on discovering and implementing innovative technologies that can mitigate the risks to others posed by convicted VNCs[265] whose behavior is predictably dangerous.[266] Contrary to popular mythology, comfortable conditions of confinement do not mean that confinement loses its deterrent effects. Norway probably has the cushiest prisons in the world[267] but one of the lowest rates of recidivism, vastly lower than the United States (20% vs. 76.6%).[268] This may not prove that comfortable confinement reduces reoffending, but it surely is evidence

264. By contrast, infliction of gratuitous hardship on prisoner and their families sometimes seems to be a deliberate feature of the American justice system. *See, e.g.*, Lorelei Laird*, Appeals Court Stymies Bid to Regulate High Cost of Prison Phone Calls,* A.B.A.J. ONLINE (May 2018), http://www.abajournal.com/magazine/article/regulate_price_prison_phone_calls; Shannon Sims, *The End of American Prison Visits: Jails End Face-To-Face Contact—and Families Suffer*, THE GUARDIAN (Dec 9. 2017), https://www.theguardian.com/us-news/2017/dec/09/skype-for-jailed-video-calls-prisons-replace-in-person-visits; *see also* Judith Miller's grim description of life for *civil* detainees in the Alexandria Detention Center, surely far from one of the worst, recounted in her book THE STORY: A REPORTER'S JOURNEY (2015), quoted at length in Darren Samuelsohn, *Prison Jumpsuit and Mystery Meats: Inside Manafort's New Jail Experience*, POLITICO (July 12, 2018, 6:09 PM), https://www.politico.com/story/2018/07/12/manafort-jail-conditions-northern-virginia-mueller-probe-716841available. No one who sees value or merit in the decent treatment of others can help but be appalled by the deliberate and calculated unpleasantness that the detention center cast into every aspect of the daily lives of the people in its charge (food, bedding, lighting, enforced loneliness and close confinement) and yet, it is a facility for people *who have not even been convicted of a crime.* One has to ask the question: What is the purpose of prison? To keep people secure (public safety) or to keep people uncomfortable.

265. *See supra* Part II.B. Note the word "convicted." I do not believe that, on the current state of the science, the prediction of criminal acts by people never before convicted is anywhere near accurate enough to justify coercive treatment.

266. I. Bennett Capers, *Techno-Policing*, 15 OHIO ST. J. CRIM. L. 495.

267. *Anders Breivik: Just How Cushy are Norwegian Prisons?* BBC NEWS (March 16, 2016), http://www.bbc.com/news/magazine-35813470 [hereinafter *Anders Breivik*]; Christina Sterbenz, *Why Norway's Prison System is So Successful*, BUSINESS INSIDER (Dec. 11, 2014, 1:31 PM), https://www.businessinsider.com/why-norways-prison-system-is-so-successful-2014-12; Erwin James, *Bastoy: The Norwegian Prison That Work*s, THE GUARDIAN (Sept. 4, 2013, 11:32 AM), https://www.theguardian.com/society/2013/sep/04/bastoy-norwegian-prison-works; Norwegian Ministry of Justice & Police, PUNISHMENT THAT WORKS (2008); NORWEGIAN CORRECTIONAL SYST. OPERATIONS STRATEGY, http://www.kriminalomsorgen.no/getfile.php/2766216.823.fvprryqpxf/Operations+Strategy+2014-2018.pdf.

268. *Anders Breivik*, *supra* note 267 ("According to the Directorate of Norwegian Correctional Service, prison should be a restriction of liberty, but nothing more"). *Cf. Recidivism,* Office of Justice Programs, National Institute of Justice, https://www.nij.gov/topics/corrections/recidivism/pages/welcome.aspx. It has been recently argued that recidivism "is not a comprehensive measure of success for criminal justice in general or for community corrections specifically." Jeffrey A. Butts & Vincent Schiraldi, *Recidivism Reconsidered: Preserving the Community Justice Mission of Community Corrections* 2, Harvard Kennedy School (Mar. 2018), https://www.hks.harvard.edu/sites/default/files/centers/wiener/programs/pcj/files/recidivism_reconsidered.pdf. Perhaps. But they are a good measure of the effects of correctional institutions on rates of victimization by prior offenders, and anyone who wants to see the rates of crime victimization go down must surely be interested in the effects of criminal justice processes on recidivism.

that it does not *increase* it.[269] We should strive to make our prisons more like Norway's, and less like North Korea's.

Even apart from the findings of neuroscience, however, there is ample reason to suspect that our nation needs a fundamental rethinking of its criminal justice system. The nation has literally become "addicted to incarceration."[270] With a quarter of American adults already having a criminal record and 1,000,000 million felony convictions per year (one every 30 seconds),[271] we are well on our way to becoming a nation of ex-cons.[272] As a result of burgeoning "collateral consequences" and the lifetime legal disabilities they impose, we are creating within our society a huge new class of second-class citizens who are shunned by employers[273] and deprived by the law[274] of the normal benefits of citizenship. That it would not be a good thing for the country to become a legally divided two-class society, with a huge underclass having an inferior set of civic rights and benefits, is hardly a matter for debate. One does not make a nation stronger by making its citizens weaker, economically and otherwise. A nation does not strengthen the well-being and dignity of its people as a whole by purposely inflicting major suffering and deprivation on an enormous and growing minority.

There are, no doubt, many reasons for America's criminal justice pathologies, but surely important among them is a widely shared commitment to the belief that blame, accountability and just deserts are immutable facts about the world rather than the social constructs that they are. To the extent that accountability and just deserts are thought proper because people's conduct is controlled by their intentions, reasons and other mental states, the mental causation hypothesis (that is embedded in

---

269. *See generally*, Rebecca Beyer, *Model Prisons*, A.B.A. J. (May. 2018), http://www.abajournal.com/magazine/article/regulate_price_prison_phone_calls; Chantal Da Silva, *New York Prisons Impose 'Draconian' Rules Limiting Books Inmates Can Read To 'Sex Novels, Bibles And Coloring Books,'* NEWSWEEK (Jan. 9, 2018), http://www.newsweek.com/new-york-prisons-impose-draconinan-rules-limiting-books-inmates-can-read-sex-775708.

270. Mark W. Bennett, *Addicted to Incarceration: A Federal Judge Reveals Shocking Truths About Federal Sentencing and Fleeting Hopes for Reform*, 87 UMKC L. REV. 3 (2018).

271. Humbach, *supra* note 226, at 606.

272. *Id.* at 605–09.

273. *Id.* at 608–09; *see* Kai Wright, *Boxed in: How a Criminal Record Keeps Americans Jobless for Life*, THE NATION (Nov. 25, 2013), http://www.thenation.com/article/177017/boxed-how-criminal-record-keeps-you-unemployed-life#.

274. The growing assortment of law-prescribed "collateral consequences" to conviction now number in the tens of thousands. *ABA Voices Concerns About the Impact of Over-Criminalization of U.S. Laws*, ABA J. (Dec. 1, 2014), http://www.abajournal.com/mobile/mag_article/aba_voices_concerns_about_the_impact_of_over_criminalization_of_us_laws (referencing National Inventory of the Collateral Consequences of Conviction, http://www.abacollateralconsequences.org (last visited Mar. 23, 2015)). *See generally* Jenny Roberts, *Why Misdemeanors Matter*, 45 U. C. DAVIS L. REV. 277 (2011).

the neuronal representation of reality of so many) is a core factor in sustaining the present exorbitance of convictions and imprisonment. To the extent, therefore that the findings of neuroscience can offer path to extirpate neuronally-embedded "beliefs" about mental causation and mitigate their contribution to criminal justice decision-making, the criminal justice system will be able to shake off its fixation on dealing with "guilt" and fulfill instead its socially indispensable role of protecting the people.

CONCLUSION

Mental causation plays an important part in the logic that justifies purposely inflicting hardship and deprivation on human beings. That being so, the epistemic case for its actual existence should be very strong. It is not. Indeed, the case is practically non-existent and, at best, it very far from being the best explanation of the totality of evidence and data that we have today as to the causes of human behavior. It matters when government purposely inflicts suffering on people who do not "deserve" it, and if mental causation is the basis for just deserts, then the doubt that neuroscience throws on mental causation is a cloud on the entire criminal justice system.

No one questions that coercive measures are sometimes needed for the sake of public safety, but the occasions and nature of these measures would be entirely different if their basis could be purged of vestigial emotions like blame and the urge to retribution. Every society has crime, but only ours has managed to attach the lifetime-label of "criminal" and "ex-convict" to 25% of its adult population. If that is not a sign and warning that something is seriously amiss in the nation's social health, it is hard to know what would be.