CRIMINAL ACTS AND BASIC MORAL EQUALITY

JOHN A. HUMBACH[*]

ABSTRACT

Modern criminal justice presupposes that persons are not morally equal. On the contrary, those who do wrong are viewed by the law as less worthy of respect, concern and decent treatment: Offenders, it is said, "deserve" to suffer for their misdeeds. Yet, there is scant logical or empirical basis for the law's supposition that offenders are morally inferior. The usual reasoning is that persons who intentionally or knowingly do wrong are the authors and initiators of their acts and, as such, are morally responsible for them. But this reasoning rests on the assumption that a person's mental states, such as intentions, can cause physical effects (bodily movements)—a factual assumption that is at odds with the evidence of neuroscience and whose only empirical support rests on a fallacious logical inference (post hoc ergo propter hoc). *There is, in fact, no evidence that mental states like intentions have anything to do with causing the bodily movements that constitute behavior. Nonetheless, the mental-cause basis for moral responsibility, though it rests on a false factual inference, has enormous implications for criminal justice policy.*

While society must obviously protect itself from dangerous people, it does not have to torment them. The imperative to punish, a dominant theme of criminal justice policy, is not supported by evidence or logic, and it violates basic moral equality.

[*] John A. Humbach, Professor of Law, Elisabeth Haub School of Law at Pace University; Ohio State University College of Law Class of 1966; B.A. in Economics, Miami University Class of 1963.

TABLE OF CONTENTS

INTRODUCTION

Is every person morally equal, or do some have less moral worth than others? Is every person equally morally deserving, or are some persons morally privileged to exploit others or use them as means—for example, to deter? We often hear that everyone is "born" or "created" equal.[1] But is this just an empty shibboleth about infants at birth, or does it describe an enduring truth about the human condition, an affirmation that basic moral equality is the indelible lifetime right of every human being?[2]

Modern criminal justice practices presuppose that people are *not* morally equal, that a person's deeds, good and bad, can affect the person's moral

---

1. G.A. Res. 217 (I) A, *See* Universal Declaration of Human Rights (Dec. 10, 1948) ("All human beings are born free and equal in dignity and rights"); *see also* THE DECLARATION OF INDEPENDENCE (U.S. 1776) ("All men are created equal"). *Cf.* DÉCLARATION DES DROITS DE L'HOMME ET DU CITOYEN DE 1789 ("Tous les Hommes naissent et demeurent libres et égaux en droits..." ("All men are born *and remain* free and equal in rights") (emphasis added)).

2. The precise contours of basic moral equality are not settled, *see generally* Stefan Gosepath, *Moral Equality*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY pt. 2.3 (Mar. 27, 2001), but the equality I have in mind for present purposes is that which flows from the fact that there is no obvious argument or evidence that some persons are morally privileged over others, to oppress others or otherwise to take advantage of others. That is to say, irrespective of the personal differences that inevitably exist in terms of natural endowments, economic accumulations or sociopolitical power, no person is morally privileged to treat any other person as an object (objectification) or to compel any person to serve as a means to further the ends of any other person, *see* IMMANUEL KANT, THE CRITIQUE OF PRACTICAL REASON 167 (Werner S. Pluhar, trans. 2002) (1785), and no one is privileged to encroach on the interests of any other person except insofar as doing so is inseparable from measures to prevent that person from encroaching on others or to effect timely restitution of enrichment gained in such encroachments. *See* ARTHUR SCHOPENHAUER, THE WORLD AS WILL AND REPRESENTATION 339-42, (E.F.J. Payne trans. 1968) (1859); ARTHUR SCHOPENHAUER, ON THE BASIS OF MORALITY 154 (Eric F.J. Payne, trans. 1995). *See also* Yitzhak Benbaji, *Welfare and Freedom: Towards a Semi-Kantian Theory of Private Law*, 39 L. & PHIL. 473 (2020) ("no person is subordinate or superior to another person"). For a non-Western precursor of ancient lineage, see the concept of universal caring (equal concern) advocated by Mozi (墨子), born c. 470 BC, and his followers see MOZI, THE ESSENTIAL MOZI: ETHICAL, POLITICAL, AND DIALECTICAL WRITINGS 51-57 (Chris Fraser, trans. 2020) ("[V]iew others' selves as you view yourself."). *See also* Chris Fraser, *Mohism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY pt. 7 (Oct. 21, 2002); Chris Fraser, *Mohist Canons*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY pt. 3 (Sept. 13, 2005). *See also infra* text accompanying notes 126-35.

In specific relation to the criminal law, basic moral equality implies that entities invested with the unique powers of government must be impartial, showing equal concern and respect (albeit not necessarily identical treatment) to every person and, specifically, regarding no one as "deserving" of hardship or deprivation at the hands of the state or otherwise. *See* RONALD DWORKIN, TAKING RIGHTS SERIOUSLY 219, 330 (1977); RONALD DWORKIN, SOVEREIGN VIRTUE: THE THEORY AND PRACTICE OF EQUALITY 405, 411 (2000) ("equally worthy of concern and respect", "equal concern and respect").

Note that basic moral equality does not necessarily entail economic equality, though is it highly relevant to economic equity, nor does it necessarily entail power equality, though it has much to say about the limits on how power should be exercised. While economic and power equality are abstractly desirable, the apparent reality is that "[i]nequality of opportunity and outcome will result from any conceivable social order as the result of innate differences in temperament and abilities, early environments, and other variables over which developing children have no control." Stephen Morse, *What Do We Owe Each Other?: An Essay on Law and Society*, L.A. REV. BOOKS (Oct. 28, 2020), https://lareviewofbooks.org/article/what-do-we-owe-each-other-an-essay-on-law-and-society/ [https://perma.cc/FD25-K9B2]. The risk of exploitation and servitude is great if economic and power equalities are sought by coercive measures that are inconsistent with basic moral equality.

worth or deservedness.[3] The law takes it for granted that those who do wrong "deserve" to suffer for their misdeeds. This is an assumption that not only underpins the retributive justification for punishment; it is one that is tacitly presupposed by utilitarian justifications as well.[4] It has, as such, enormous practical relevance in assessing the justness and morality of criminal law.

Criminal justice practices would be very different if we did not assume that wrongdoers are morally inferior in their entitlement to respect, concern, and decent treatment. Not only do the utilitarian and retributive justifications for punishment depend on this assumption, but, if we did not make it, we would lose an important excuse for the deplorable conditions in which today's prisoners are normally held.[5] But to say that persons' deeds affect (or reveal) their moral worth or deserts is, as a practical matter, to say that human beings do not have basic moral equality. It is to say that, at any

---

3. *See* Paul H. Robinson, *Are We Responsible for Who We Are? The Challenge for Criminal Law Theory in the Defenses of Coercive Indoctrination and "Rotten Social Background*," 2 ALA. CIV. RTS. & CIV. LIBERTIES L. REV. 53, 62-65 & 74-76 (2011). *See also* Paul H. Robinson & Lindsay Holcomb, *Indoctrination and Social Influence as a Defense to Crime: Are We Responsible for Who We Are?*, 85 MO. L. REV. 739 (2020); MATT 7:16 ("Ye shall know them by their fruits"); J.K. ROWLING, HARRY POTTER AND THE CHAMBER OF SECRETS 333 (1998) ("It is our choices, Harry, that show what we truly are….").

4. The premise of the *retribution* rationale for punishment is of course that criminals deserve to be punished. *See* MICHAEL MOORE, PLACING BLAME: A THEORY OF THE CRIMINAL LAW 87-91 (2010). Although the *utilitarian* rationales for punishment (deterrence, incapacitation, rehabilitation) do not rely explicitly on deserts for their justification, it is hard to avoid the conclusion that they could not stand without an underlying tacit assumption that offenders deserve to be punished. For example, most utilitarians would almost certainly balk at deliberately imprisoning innocents (e.g., the children of offenders) no matter how much deterrence or other social benefits such punishments might yield. And most utilitarians would agree, presumably, that "fair trials" are important because only the factually guilty should be punished. For my further discussion of these points, see John Humbach, *Neuroscience, Justice and the "Mental Causation*" Fallacy, 11 WASH. UNIV. JURIS. REV. 191, 246-47 (2019) (hereinafter "*Mental Cause Fallacy*").

5. For some descriptions of conditions of confinement, see e.g., Michael B. Mushlin, *Our Jail and Prison Shame, Coast to Coast*, N.Y. DAILY NEWS (Oct. 6, 2021), https://www.nydailynews.com/opinion/ny-oped-our-prison-shame-coast-to-coast-20211006-y2itjuns2nchbc74qcecso7gtq-story.html [https://perma.cc/9BE3-G5QX]; Debra Cassens Weiss, *Federal Judge Complains That New York's Federal Detention Facilities 'Are Run by Morons'*, A.B.A. J. (May 10, 2021), https://www.abajournal.com/news/article/federal-judge-complains-that-new-yorks-federal-detention [https://perma.cc/3RLV-FCQ9]-facilities-are-run-by-morons ("facilities . . . were so cold that her tears froze on her face, moldy food, sandwiches that were so frozen they hurt her teeth, and filthy conditions," e.g., mice, ankle-deep feces floods); Daniel Genis, *Where's Harvey Weinstein Headed? It's a Place I Know*, QUILLETTE (Apr. 4, 2020), https://quillette.com/2020/04/04/wheres-harvey-weinstein-headed-its-a-place-i-know/ [https://perma.cc/5ZDU-NLBV]; Michael Rothenberg, *The Federal Prisoner Transit System—aka "Diesel Therapy"—Is Hell*, THE MARSHALL PROJECT (Aug. 15, 2019), https://www.themarshallproject.org/2019/08/15/the-federal-prisoner-transit-system-aka-diesel-therapy-is-hell [https://perma.cc/65BP-ZRP2]. *See also* Jacob Bronsther, *Long-Term Incarceration and the Moral Limits of Punishment*, 41 CARDOZO L. REV. 236 9 (2020) (making the case that excessively long punishments are "ruining" people's lives). Prisoners are even cheated economically by systematic price-gouging on their telephone calls to family and others, and almost nobody in authority seems to see a problem with these practices, which have only begun to be remedied. *See* Matt Reynolds, *FCC Approves Plan to Make Some Phone Calls Cheaper for Inmates and Their Families*, A.B.A. J. (May 21, 2021), https://www.abajournal.com/news/article/fcc-curbs-out-of-state-call-rates-in-prisons [https://perma.cc/3ZBL-YASJ]. One can only imagine what casual cruelties are less visibly inflicted.

given time, some people are morally superior and privileged to inflict harm on their moral inferiors. Such an idea is strikingly out of line with today's professed commitment to at least minimal levels of equality. Yet, the reasons for thinking that bad deeds can affect (or reveal) basic moral worth or deserts are rarely discussed.

I. A CASE AGAINST BASIC MORAL EQUALITY (STEINHOFF AND KAGAN)

Uwe Steinhoff proposes a hypothetical that he thinks demonstrates what "pretty much everybody apart from egalitarians" already knows, namely, that people are "simply not moral equals."[6] The gist of Steinhoff's hypothetical is this: A person in a small boat comes on two distressed swimmers, Dr. Albert Schweitzer[7] and Adolf Hitler, both drowning in the sea. Only one of them can be saved. Which one should it be?[8] Steinhoff concludes that the altruistic Dr. Schweitzer is the one to save because he has greater moral worth. Using analogous scenarios, Shelly Kagan concludes that some people are more morally *deserving* than others.[9] He presents a scenario in which Amos and Boris are gravely injured in an explosion. Boris was at fault in causing the blast. If only one of the two can be saved, which one should it be?[10] Kagan's answer is, predictably, Amos based on the crisply worded aphorism "*fault forfeits first.*"[11] In sum, Steinhoff and Kagan maintain that a person's moral worth or deservedness is not fixed for life but can vary over time. Good behavior increases a person's moral deservedness and bad behavior reduces it.[12] Basic moral equality, if it ever

---

6. UWE STEINHOFF, WALDRON ON THE "BASIC EQUALITY" OF HITLER AND SCHWEITZER: A BRIEF REFUTATION 2 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3401698 [https://perma.cc/2T3U-VGR9].

7. Albert Schweitzer was an iconically renowned physician, musician, and humanitarian.

8. STEINHOFF, *supra* note 6. One may justly question the value of bringing personalities like Schweitzer and Hitler into intuition-pumping moral hypotheticals. Though extreme-case scenarios may yield moral intuitions that are crystalline in their clarity, it is hard to be sure that such intuitions reliably reflect the moral sentiments that would be evoked in real life situations. DANIEL C. DENNETT, FREEDOM EVOLVES 182 (2003) ("Thought experiments that stipulate such extreme—and extremely unrealistic—conditions are notoriously likely to beguile the philosopher's imagination…."). *See also* James Wilson, *The Trolley Problem*, AEON (May 28, 2020), https://aeon.co/essays/what-is-the-problem-with-ethical-trolley-problems [https://perma.cc/CP5T-SWTM], for a good discussion of reasons why it is problematic to rely on philosophical thought experiments for information about ethical norms.

9. SHELLY KAGAN, THE GEOMETRY OF DESERT 23-44 (2012).

10. *Id.*

11. *Id.* It is, to put it mildly, far from clear how this "forfeiture" works. It does not seem plausible that persons forfeit moral rights just because others *say* they do or use the magical word "forfeit." The law can create legal rights and provide for their forfeiture. But by what authority can one who did not create a moral right make the right forfeitable?

12. Douglas Husak cites Kagan in order to make a similar point. Douglas Husak, WHAT DO CRIMINALS DESERVE? 5-6 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2788152 [https://perma.cc/9RHS-5W85]. Though Husak's analysis is different from Kagan's, he does not appear to doubt that Kagan is correct in his claim that "some people are more morally deserving than others." The problem with Husak's analysis is that it appears to rely, for its key analytical move, on the naturalistic fallacy—arguing that because people naturally have negative reactions to others' wrongful

exists at all, is a highly ephemeral affair.

Steinhoff and Kagan both seem satisfied that their scenarios are adequate to demonstrate that people are morally unequal. But they do not. For one thing, the scenarios both assume that readers will focus on the moral worth or desert of the persons in question as the operative reason for deciding whom to rescue. But because the scenarios do not adequately exclude other potential influences on the moral intuitions they evoke, those intuitions may not be reliable gauges of moral worth or deserts. As Jeremy Waldron has pointed out (in defending the position that all persons *do* have equal moral worth), there are other moral principles that "permit all sorts of differentiation" in deciding which of Steinhoff's swimmers should be saved.[13] The same can be said of the Amos and Boris scenarios. As examples of other principles, Waldron mentions denunciation, punishment, and targeted killing, all of which (he assumes) could also serve as morally permissible bases for differentiation for deciding whom to save.[14] In other words, quite apart from considerations of moral worth or deserts, one might decide to save Dr. Schweitzer and leave the genocidal dictator to his fate (or save Amos and forsake Boris) in order to denounce, to punish, or simply to neutralize a threat—all being (according to Waldron) morally permissible bases for differentiation.

Steinhoff dismisses Waldron's counterexamples on the ground that the moral differentiations they authorize are factually distinguishable from the decision to "let die" in his swimmer story. However, even if Waldron's counterexamples are distinguishable (a debatable point[15]), it would not undermine his larger point. Waldron's larger concern (which seems beyond dispute) is that, in real-life situations, with complex factual contexts, there are likely to be plenty of reasons, other than basic moral worth or deserts, to want to save someone like Dr. Schweitzer rather than someone like Hitler.[16]

---

conduct, the wrongdoers deserve the negative treatment (punishment) they receive as a result of those reactions.

13. JEREMY WALDRON, ONE ANOTHER'S EQUALS: THE BASICS OF HUMAN EQUALITY 150 (2017).

14. *Id.*

15. It is, in my view, doubtful that the facts of Waldron's counterexamples are *relevantly* distinguishable: In a nutshell, Steinhoff's argument for distinguishability is that "letting" a person die is not the same as purposely killing him, denouncing him or harming him with "intention" to punish. It seems to me, however, that Waldron's three counterexamples are all fair analogies in that they all share with Steinhoff's hypothetical the relevant common feature of a disposition to countenance prima facie immoral treatment of another which, in each case, is justifiable by some countervailing principle. Waldron's counterexamples therefore *do* support Waldron's position that, owing to the controlling application of other possible moral principles, Steinhoff's hypothetical does not unambiguously make the case that people do not have basic moral worth.

16. For example, as Professor Peter Kostant once pointed out to me, it's always good to have a doctor in a lifeboat. Just be clear, I would for my own part intuitively prefer to save someone like Schweitzer (or Amos), but I would not draw the conclusions from the preference that Steinhoff and Kagan propose.

II. USING PAST DEEDS TO ASSESS MORAL WORTH OR DESERTS

There is, however, a more fundamental reason why the Steinhoff/Kagan scenarios do not show what they are meant to show: Although both rely on a person's past deeds as criteria to assess the person's moral worth or deserts, neither offers any explanation or reasoning to justify that reliance. An important logical step seems to be missing.

It is not, after all, simply self-evident that "fault forfeits first" or that the things people do have any connection to their moral worth or deserts. Obviously, some deeds are more morally worthy than others,[17] but it does not follow that some *persons* are. Yet, both Steinhoff and Kagan seem to take it for granted that a crucial connection or nexus exists between the moral quality of deeds and the moral worthiness of those who do them. Steinhoff, for example, jumps straight to the conclusion that the conduct of the "life-saving philanthropist" Dr. Schweitzer means he has acquired or evinced[18] greater moral worth than genocidal Adolf.[19] As for Kagan, he seems to simply stipulate that *fault forfeits first* and, based on that, declares that bad-deed Boris is morally less deserving than Amos.[20] What is missing in both cases is any argument as to *why* people's deeds are proper criteria for assessing moral worth or deserts, criteria that are less arbitrary than, say, hair color, math skills or shoe size. Both Steinhoff and Kagan leave us to wonder why the simple fact of being human does not suffice *in itself* to establish and fix the basic moral worth and deserts of every person.

If there is a connection between a person's deeds and her moral worth or deserts, it almost certainly rests on the assumption that human beings are the origins and authors of their acts, *viz*. that persons are "agents" who initiate their own bodily movements.[21] Only by connecting the origination of a person's deeds to the person who does them can moral responsibility be attached. What it takes to make a person an "agent" is anything but settled,[22] but it seems clear that mere *physical* involvement in a causal chain of harm is not enough. Consider, for example, a crowd viewing a regatta from a pier when a sudden gust of wind throws a hapless spectator off-balance so she bumps another into the roiling sea. Would we assess the hapless spectator's moral worth based on that? Probably not. Because of the

17. *See infra* text accompanying notes 126-35.

18. Neither Steinhoff nor Kagan makes clear whether he thinks that bad deeds *make* a person less morally worthy (or deserving) or, in the alternative, that bad deeds *manifest* an already-existing moral inferiority, as in "Ye shall know them by their fruits." MATT 7:16. For present purposes, however, the distinction makes no difference. Both Steinhoff and Kagan agree that bad deeds are relevant to assessments of moral worth, or deserts, and that is the position this paper takes issue with.

19. STEINHOFF, *supra* note 6, para. 1. 1.

20. KAGAN, *supra* note 9, at 26-25.

21. *See generally* Marcus Schlosser, *Agency* 2.2, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Aug. 10, 2015).

22. *Id.*, introduction at 1.

obvious role played by the wind in causing the harm,[23] we can easily see that the hapless spectator was not the origin or author of bump; she was merely a *conduit* for harm-producing causal forces that came from elsewhere. The situation is not much different in principle from that of a tree that conducts a bolt of lightning down so it strikes someone sheltering beneath. The tree is merely a conduit for the harm, not the initiator. If it is arbitrary to morally judge people based on acts that are not their own,[24] it is equally arbitrary to place blame on those who are mere conduits for causal chains of events that come into them from outside. Thus, when Steinhoff and Kagan take it for granted that a person's deeds affect or reveal the person's moral worth or deserts, they must be assuming, albeit tacitly, that human bodily movements are not originated externally (like the bump off the pier) but are initiated by some function or attribute of the person who performs them.

There is, however, a problem with the assumption that persons themselves are the authors and origins of their own bodily movements. Physical motions are not known to just occur out of nowhere.[25] Rather, according to the laws of physics, a physical motion can only occur if there are prior causal events generating forces that suffice to make the motion occur.[26] And these physical laws do not only apply to inanimate objects. They also apply to movements of the human body and to the motions of the billions of particles that make its muscles and neurons work (ions, neurotransmitters, etc.).[27] In other words, *as a purely physical matter*, the complex chains of neuronal events that come together to produce human

---

23. Note that the text says the wind played a causal role, not that it was "the" cause of the harm. It is taken for granted that nearly every event has more than one cause—and that, typically, the causes of an event are countless going back in time (*causa causae est causa causati*).

24. *See* James W. Moore, *What Is the Sense of Agency and Why Does it Matter?*, 7 FRONTIERS PSYCH. 7 (Aug. 29, 2016) ("[F]or most people it only makes sense to hold someone responsible for their actions if they are freely in control of them"); OLIVER WENDELL HOLMES, THE COMMON LAW 54 (1881) ("[I]t is felt to be impolitic and unjust to make a man answerable for harm, unless he might have chosen otherwise").

25. Apart from random quantum events which, for these purposes, can be ignored. Quantum mechanics presents no substantial exceptions to classical physics principles in the application of those principles to ordinary human physiology or to the social interactions that are the concern of the law. *See infra* note 26. The quantum effects are always there, of course, but in human-sized systems they lose their quantum coherence, which makes them virtually impossible to detect without specialized equipment. *See* CARLO ROVELLI, HELGOLAND: MAKING SENSE OF THE QUANTUM REVOLUTION 52-53, 210 n. 39 (Erika Segre & Simon Carneli, trans. 2021); Scott Aaronson, *Quantum Randomness*, AM. SCIENTIST, https://www.americanscientist.org/article/quantum-randomness [https://perma.cc/U4XV-7XM9].

26. The classic formula is F=ma (a.k.a. Newton's Second Law of Motion), which tells us how much force is needed to increase the velocity of a given mass by a given amount within a given time. Another way stating this concept is, simply: Physical objects do not move around except in accordance with the laws of physics. This is sometimes referred to as the "causal closure" of the physical domain. *See* JAEGWON KIM, PHYSICALISM, OR SOMETHING NEAR ENOUGH 15, 36–45 (2005).

27. *See* Sean M. Carroll, *Consciousness and the Laws of Physics*, 28 J. CONSCIOUSNESS STUD. 16 (2021) ("physical events have purely physical causes . . ., at least in the regime relevant to human life"), https://philpeople.org/profiles/sean-m-carroll [https://perma.cc/J764-8BFL]. *See generally* sources cited *infra* note 29 and 47 for information on how bodily movements are neuronally produced.

bodily movements and, hence, behavior could never originate spontaneously *inside* the person.[28] In order for the chains of neuronal events to get started *inside* the person there would first have to be physical forces acting on the body from *outside*—either recently or in the past[29]—mostly, one may presume, in the form of sights and sounds coming in through the senses.[30] In other words, if human bodily movements are a *purely physical matter*, then we are, in every move we make, merely conduits for chains of physical events initiated elsewhere—like the gust at the regatta. As a purely physical matter, persons no more have "control" of what they do than a steering wheel controls which way the car goes.[31]

Because Steinhoff and Kagan quite evidently think that persons *do* have agential control over their actions, they must be assuming that the production of human behavior and choices is *not* a "purely physical matter" (or is not, at least, dictated in accordance with physical laws). They must be assuming instead that there is something about the person, something *not* subject to physical laws, that can initiate the person's actions and, by virtue of such initiation, make the person morally responsible for them.[32] For it is

28. While random quantum events inside the body could, in theory, initiate random bodily movements, such movements would fall more into the category of spasms rather than "behavior." *See supra* note 25.

29. *See* ROBERT M. SAPOLSKY, BEHAVE: THE BIOLOGY OF HUMANS AT OUR BEST AND WORST (2017), who, over the course of 800+ pages, provides a meticulously detailed description of the kinds of causes and factors that determine what a person does, beginning seconds or less before actions occur and stretching far back through time.

30. Of course, the causative "sights" enter the body, not as complete visual scenes, but as individual quanta of light that enter the retinal cells and trigger them to "fire." What the eye itself "sees" is only a scattering of dots. Much the same can be said of sounds, which enter the body as successions of pressure variations in the ears. Both the dots and the pressure variations have to be interpreted internally (*computationally*, if it's all just physical) to infer the external scene that produced them. Based on these inferences, together with information from prior perceptions that are registered in memory, something in the person must then make a determination (again computationally, if it's all physical) whether to prompt movements in response. The physical paths that these forces take through the body-as-conduit, from sensory input sites to contracting muscles, are (as will be discussed below) essentially neuronal. *See infra* text accompanying notes 56-67. But the point for present purposes is this: As a purely physical matter, bodily movements (behavior) cannot arise inside the person "out of nowhere," so to speak, but can only occur if external forces, such as quanta of light, phonons of sound, tactile pressures, etc., enter the body and trigger internal chains of causation that produce them.

31. Essentially just like a steering wheel, a person's physical morphology and physiology are mechanistic conduits that conduct the chains of causal forces that act upon the body over time. It can be said, of course, that "the person" plays a causal role in controlling her behavior in the sense that her it is the person's own physical morphology and physiology are proximally operative as conduits in determining what she does. But *if it's all just physical*, these proximal intermediating links in the causal chains (i.e., morphology and physiology) cannot confer the person with any kind of autonomous control as an "agent," as it were, any more than any other conduit for forces can be properly regarded as exerting control on its own. As a conduit for the chains of causal forces that produce a person's movements, the person's morphology and physiology do not "have control" except in the sense that a car's steering wheel has control—or, more complexly, in the sense that a computer keyboard controls the words displayed on the screen, or that an autopilot device controls motions of an airplane.

32. As Kagan says: "I will simply assume that there is some relevant desert basis, and that in principle, at least, we can rank people differentially in terms of this basis." KAGAN, *supra* note 9, at 6. He even declines to go so far as to assert that it is a person's deeds that make the moral difference,

only by assuming that behavior comes from something about the person herself, other than just her deterministic "physical self," that Steinhoff and Kagan can establish that the person is an "agent" rather than a "conduit," thereby supplying the connection or nexus that is needed between the person and her deeds to make the person morally responsible for them.

III. MENTAL CAUSATION

Commonly thought to supply the morally crucial connection between deeds and moral worth, are the person's mental states (such as intentions, volition, desires, or reasons) that seem to make the body move.[33] Mental states like these make most human actions seem plainly different from the bump at the regatta. And, of course, mental states are a paramount concern of the criminal law.[34] They are the law's criteria for voluntary action,

---

declaring the matter to be outside the ambit of his investigation. *Id.* at 6. However, as noted earlier, *see supra* text accompanying notes 9-12 and note 18, Kagan makes amply clear that a person's deeds figure into his assessment of moral deserts.

33. In general, the term "mental state" is used in this discussion with its ordinary everyday meaning, particularly the meaning as understood in law. *See* 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 (2007). Just to be clear, however, the term "mental state" is not used to refer to purely *physiological* states or "brain states." Rather, the reference is to experiential states that have some element of consciousness, qualia, thought, awareness or the like. In other words, the "mental causation" discussed in this article is meant to refer to the putative causal effects of *non-material* forces or presences, such as intentions, reasons, thoughts, desires, the will or the "mind."

It has sometimes been argued, in defense of mental causation, that experiential mental states may be "identical" to, "reduced" to or "properties" of underlying physiological brain states and, therefore, be able to share the physical brain's causal efficacy (identity theory, reductionism and property dualism). The law takes no explicit position on such possibilities and neither does this article. As I have discussed elsewhere, however, insofar as mental states depend on physical processes for their occurrence and content, their causal effects would have no independent moral significance beyond that which inheres in the physical processes and interactions on which they depend. *See infra* text following note102; text accompanying notes 110-112; John Humbach, *Do Criminal Minds Cause Crime? Neuroscience and the Physicalism Dilemma*, 12 WASH. UNIV. JURIS. REV. 1, 5-6, 13-25 (2019) (for a more extensive discussion). In other words, for mental causes to be relevant to moral judgments about persons based on their conduct, the occurrence and content of the causative mental states would have to be autonomous of the deterministic physical domain. *Id.*

34. *See, e.g.*, Elonis v. United States, 575 U.S. 723, 734 (2015) (stating that "the basic principle [is] that 'wrongdoing must be conscious to be criminal'" (quoting Morissette v. United States, 342 U.S. 246, 250 (1952)) 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) (providing a 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 F. MANDIBERG, *Strict Liability*, THE ENCYCLOPEDIA OF CRIMINOLOGY AND CRIMINAL JUSTICE 2222 (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 . . .

culpability, justifications, and excuses and, as such, they "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." [35] Even more importantly for present purposes, unlike the purely physical causes of behavior discussed in the previous section, a person's mental states (such as intentions and desires) do *not* appear to be driven or determined in accordance with physical laws: They do not appear to *need* sufficient prior causes. On the contrary, our streams of conscious awareness seem to wend their way freely through the labyrinths of reason and thickets of thought, quite unconstrained by the physical rules of cause and effect. They are what Donald Davidson calls "anomalous" (*lit.* not subject to laws). As Davidson has declared: "there are no strict deterministic laws on the basis of which mental events can be predicted and explained."[36] And, indeed, from ordinary subjective introspection it is easy to infer that our mental states often have no external causes. They sometimes seem to just occur "out of nowhere," originating *causa sui* as the person's own creation. They seem quintessentially our own.

Because mental-state causes of behavior (like intentions) are anomalous and often occur without apparent external causes, mental causation seems to provide a perfect answer to the physicalist claim[37] that we are, in every move we make, merely conduits rather than agents. A mental-cause connection between a person's own self-generated, "anomalous" mental states and her behavior would provide exactly the link we need to justify holding the person morally responsible for what she does and using her deeds to assess her moral worth and deserts. While Steinhoff and Kagan do not say so, it seems almost certain that both regard mental causation (by intentions, reasons, etc.) as the factor that supplies the morally crucial nexus between the things people do and their moral worth or deserts.

The trouble is this: If there were *in fact* a causal connection[38] between a person's mental states and the person's deeds, one would expect there to be at least some evidence for it. But there is not. In order for mental states to

---

intends his movements to cause [the prohibited result, and knows they will cause the prohibited result].").

35. Stephen J. Morse, *The Inevitable Mind in the Age of Neuroscience in* PHILOSOPHICAL FOUNDATIONS OF LAW AND NEUROSCIENCE 35 (Dennis Patterson & Michael S. Pardo eds., Oxford 2016).

36. Donald Davidson, *Mental Events in* ESSAYS ON ACTIONS AND EVENTS 138 (Oxford 2001) (1980). It is not surprising that "psychology has no laws" because, to have scientific laws, there have to be fungible entities (such as elementary particles or units of mass or energy) or, at least, entities that are fungible with respect to the characteristics that the laws address and that affect their operation.

37. *See supra* text accompanying notes 25-31.

38. Note the wording here, which refers to a causal connection and not merely to a causal relationship, a choice of words that is meant to highlight the irrelevance of causal theories (such as "counterfactual" or "difference-making" causation) that declare the existence of causal relations among events based on *correlations* alone. *See infra* notes 72 and 88; *see also infra* text accompanying notes 80-89.

make the body move, there would have to be some kind of *interoception* by which the brain and central nervous system (or, at least, the muscles) could detect the contents of mental states and bring about bodily movements that are responsive to them.[39] But despite years of study of the processes that make the body move,[40] not a shred of evidence has been found that such interoception exists or that there are structures in which it might occur. There is, in short, nothing whatever to show there even *is* such a thing as mental causation, either of human bodily movements or of anything else that happens in the physical domain. Everybody knows that people cannot move billiard balls or chess pieces just by simply "willing" them to move. The mind does not move molecules. There is likewise no evidence that persons can move the parts of their own bodies just by willing it either.[41]

What passes for evidence of mental causation is essentially nothing more than a blunt "just so" story. Specifically, what we know comes from simple introspection and it is this: (1) Certain mental states (e.g., intentions and desires) seem to precede and portend particular bodily movements, and (2) the movements that the mental states portend then occur, like Tuesdays follow Mondays. In other words, the whole case for mental causation consists of an uncorroborated inference from a naked correlation, with nothing to suggest that there is any actual *connection* between mental states and deeds. And it is a fallacious inference at that, —namely, the logical fallacy of *post hoc ergo propter hoc*.[42] The whole case for behavior-based

39. The brain is, of course, informed of the contents of mental states but that is only to be expected if the brain is the proximal *source* of that content. It is, of course, still one of the great unknowns where and how mental states get their content, but it seems nearly certain that they must depend on the brain, senses and nervous system to supply at least some of that content. After all, as far as anyone knows, the mind and consciousness have no sensory apparatus or detectors of their own or any other source of information about the outside world. *See infra* text accompanying notes 104-109.

40. According to a study conducted by Elsevier, 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*, ELSEVIER (2014), https://www.elsevier.com/connect/new-reportmaps-the-landscape-of-global-brain-research [https://perma.cc/T9R6-5F7P]; *see also* RICHARD PASSINGHAM, COGNITIVE NEUROSCIENCE: A VERY SHORT INTRODUCTION 3 (2016) (stating that "nearly 30,000 experiments conducted using fMRI alone."). Much of this research is described and summarized in Sapolsky, *supra* note 29.

41. That is, there is no evidence that a mere "will" to make something move, no matter how strong, could put an arm or a leg into motion, or move molecules at a synapse or ions at an axon's sodium gate. For these events occur, the evidence shows there has to be a sufficient physical cause, specifically, successions of physiological causes that trace back through the central nervous system and brain to the senses and beyond. *See supra* notes 29-31 and accompanying text and *infra* 53-56 and accompanying text for a discussion of what the evidence *does* show concerning the causes of bodily movements.

I do not, by the way, discuss Benjamin Libet's famous experiments on behavior production because, in my view, some of Libet's experimental methods led to results that are ambiguous and subject to competing interpretations on the key issues relevant to mental causation. See Benjamin Libet, *Do We Have Free Will?*, 6 J. CONSCIOUSNESS STUD. 47, 51 (1999), (explaining that "the initiation of the freely voluntary act appears to begin in the brain unconsciously, well before the person consciously knows he wants to act.").

42. Literally, "after this, therefore on account of this." The *post hoc* inference is a "false cause" fallacy that is "inherently mistaken" and not valid. *See* Leo Groarke, *Informal Logic*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2017), https://plato.stanford.edu/entries/logic-informal/ [https://perma.cc/KW4C-22YA]; Hans Hansen, *Fallacies*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY

moral inequality in worth or deserts is built, at best, on a logical fallacy.[43]

Before the advent of modern neuroscience, this lack of evidence or valid inference for mental causation was not much of a concern. The reason is not hard to see. Every psychologically normal person has a vivid subjective experience of intentions, desires, beliefs reasons, and other action-related mental states, and it is perfectly obvious that what we do is highly correlated with these mental states. A conscious intention to swat a fly typically precedes the swing of the swatter; a desire for light precedes the flick of the switch. Consistent correlations like these, and our resulting "sense of agency,"[44] demand explanations. Before modern neuroscience had documented a *physiological* explanation for the correlations and sense of agency, mental causation was the only game in town.[45] Thus, mental causation became, by default, the story people told to connect the things

---

(2015), https://plato.stanford.edu/entries/fallacies/ [https://perma.cc/G4NS-2592]; Bradley Dowden, *Fallacies*, INTERNET ENCYCLOPEDIA OF PHILOSOPHY, http://www.iep.utm.edu/fallacy/#FalseCause [https://perma.cc/8L6E-P6WZ].

43. For my lengthier examination of the weakness of the case for mental causation and of the logical fallacy that is involved, *see Mental Cause Fallacy*, *supra* note 4, at 224-34.

There is, to be sure, an abundance of philosophical discussions favorable to mental causation, but those discussions almost invariably endeavor to show only that mental causation is *possible*, not that it is a fact. *See, e.g.*, Michael S. Moore, *Contemporary Neuroscience's Epiphenomenal Challenge to Responsibility* 198 & 198-203, *in* 6 OXFORD STUDIES IN AGENCY AND RESPONSIBILITY (David Shoemaker, ed. 2019) (arguing against epiphenomenalism). By contrast, the facticity of *neuronal* causation is just about as well established by evidence as any fact of biology or physiology can be.

44. *Mental Cause Fallacy*, *supra* note 4, at 229-32. The sense of agency is the "feeling of being in the driving seat when it comes to our actions." See Moore, *supra* note 24 at 1. It is however facile to assume that the so-called "sense of agency" provides evidence of mental (as opposed to physical) causation. *Mental Cause Fallacy*, *supra* note 4, at 229-32. There are a couple of reasons: First, the neuroscience evidence shows that the brain's computation of the sense of agency is most likely based on the brain's detection of correlations among physiological events associated with the generation of motor signals but with no indication that the brain detects mental causation or mind/body interactions as such. *See* citations and discussion in *Mental Cause Fallacy*, *supra* note 4, at 230, n. 186-88. Second, although your sense of agency does tell you that "you" are in the driver's seat and that "you" (rather than something else) made such-and-such happen, it does not distinguish which "part" of you (mind or physiology), made it happen. You can "feel" your sense of agency informing you that you made your body move, but it is pure speculation to conclude that your mind rather than your physiology is what made the movement occur. Recent research tends to confirm this conclusion. *See* Keisuke Suzuki et al., *Intentional Binding Without Intentional Action*, PSYCH. SCI. (2019), https://journals.sagepub.com/eprint/KESEFKVU6FXBMTX7XJMP/full [https://perma.cc/FN7A-HRVB]. *See* my further discussion at *Mental Cause Fallacy*, *supra* note 4, at 229-32.

45. Well, just about the only game, anyway. Thinkers such as Malebranche and Leibnitz offered hypotheses that are even *harder* to accept. *See* Tad M. Schmaltz, *Nicolas Malebranch*e, *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 [https://perma.cc/9LAS-STXK]("God . . . brings it about that our sensations and volitions are correlated with motions in our body."). Leibnitz provided another competing explanation based on an alleged 'pre-established harmony' of mind and body, rather like the clocks that were synchronized by the shopkeeper in the morning. *See* JAEGWON KIM, PHILOSOPHY OF MIND 171 (2011); *Gottfried Wilhelm Leibniz* § 4.4 (2013), https://plato.stanford.edu/entries/leibniz/#PreEstHa [https://perma.cc/9P8Q-2B9D]. In the days of witchcraft, spooks and other numinous forces, mental causation provided a more parsimonious plausible explanation than Malebranche or Leibnitz.

they did with the intentions, desires, volitions, and other mental states they experienced.

Today, however, the traditional mental-causation explanation has competition. Modern neuroscience research presents a very different story, a purely physiological story,[46] to connect our thoughts and our actions.[47] It is a story that is, moreover, grounded in material reality and substantiated by literally millions of experiments.[48] Unlike the hypothesis of mental-state causation, the neuroscience explanation has the virtue of explaining the production of human behavior in a way that is fully consistent with the physical laws that apply throughout the Universe.[49]

According to the neuroscience account, every bodily movement (and, hence, all behavior) is the product of ordinary physiological functioning.[50] Human beings, like other organisms with brains, are complex adaptive systems that physiologically process information acquired by the senses and somatic interoception to produce, via sequences of minute synaptic firings, organized cascades of motor impulses that trigger coordinated contractions of countless sarcomere (muscle) fibers.[51] Consonant with this mechanismic,

---

46. Much of this research is described and summarized in Sapolsky, *supra* note 29; and PASSINGHAM, *supra* note 40.

47. *See infra* following paragraph in text. My understanding of the neuroscience description of human behavior is based primarily on the following (in addition to numerous articles): Sapolsky, *supra* note 29; PASSINGHAM, *supra* note 40; BRYAN KOLB & IAN Q. WINSHAW, FUNDAMENTALS OF HUMAN BRAIN AND BEHAVIOR (4th ed., 2012); ANTHONY DAMASIO, DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN (2005); PATRICIA S. CHURCHLAND & TERRENCE J. SEJNOWSKI, THE COMPUTATIONAL BRAIN (1992) (information processing across biological neural networks); IRA B, BLACK, INFORMATION IN THE BRAIN (1991) (focusing on molecular level); DAVID H. HUBEL, EYE, BRAIN AND VISION (1988) (brain information processing with emphasis on visual information); PATRICIA S. CHURCHLAND, NEUROPHILOSOPHY: TOWARD A UNIFIED SCIENCE OF THE MIND/BRAIN (1986) (comprehensive essay relating neurophysiological findings to the perennial "mind/body" problem); JEAN-PIERRE CHANGEUX, NEURONAL MAN: THE BIOLOGY OF THE MIND (1985) (general introduction to brain structure and its functioning in information processing); *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); and DANIEL 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).

48. *See* DAVIDSON, *supra* note 36.

49. As discussed *infra* note 79, this article's commitment to physicalism (or metaphysical materialism) is strong but provisional. The truth is we do not know, and have at least some reason to doubt, the classical physics picture of reality consisting of material objects moving through space and time. *See* Rovelli, *supra* note 25, at 118-58. But even though nature may be an edifice whose foundations we will never know, that does not mean we cannot study its upper floors and the intricate relations found there and profit greatly from it. And notwithstanding certain quantum-mechanics phenomena, the prevailing interpretation of what we do know is that reality has a physical foundation. *See* 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 [https://perma.cc/7P9C-3G7X].

50. *See* Mark Hallett, *Volition: How Physiology Speaks to the Issue of Responsibility* (2011), *in* CONSCIOUS WILL AND RESPONSIBILITY 61, 65 (Walter Sinnott-Armstrong & Lynn Nadel, eds. 2011); Sapolsky, *supra* note 29, at 21-77; PASSINGHAM, *supra* note 40, at 66-81 (2016).

51. You may think you can sense your mind moving your arms and legs, but one thing is clear: That does not happen. Your arms and legs are moved by muscles, and those muscles are activated by signals arriving via motor neurons. That much is indisputable. Therefore, the only way your mind could

computational description of behavior production, the reason certain mental states are so highly correlated with actions is not that mental states cause the body to move. Rather, the reason for the correlation is that our bodily movements and the mental states are *both* brought about by the same third factor, viz., the neuronal activity that makes the body move also supplies the content of the accompanying conscious mental states.[52]

Is this neuroscience account of behavior production true? This question is, of course, a factual one, which can be resolved only by evidence, not by metaphysical reasoning alone. But there is an enormous body of evidence to substantiate the physiological (neuroscience) account[53] and only ambiguous correlation evidence, plus logically fallacious reasoning, to support its mental-cause competitor. Indeed, to think that mental states cause behavior despite the existence of a vastly better supported physical explanation is like seeing a carpenter pound a nail and then thinking, despite the evidence, there is something in the wood that sucks the nail down.[54]

In sum, if we are going to justify treating people harshly based on a belief, we should presumably want the belief to be better supported than the belief that intentions cause actions. Given the overwhelming weight of the evidence favoring the neuroscience alternative to mental causation, the justifiability of current criminal justice practices is dubious.[55]

---

move your arm or leg would be by activating neurons in the motor areas of the brain to organize the coordinated cascades of neuronal impulses that are needed to produce muscular contractions and, hence, your behavior. But you can no more move your body by mental states alone than you can putt a golf ball just by thinking about it.

52. *See* Hallett, *supra* note 50, at 66. It is, of course, famously unknown how the physical brain could supply content to mental states or consciousness, and neuroscience research has not shown that it does. It does, however, seem to be a fair surmise that the mind does not have direct "sensory" access to the outside physical world. It seems instead to be nearly certain that mental states acquire their information about external reality via the brain, central nervous system and sensory modalities of sight, hearing, etc. Happily, however, there is no need to resolve this question for purposes of the point being advanced here, viz. that it is, on the evidence, practically certain there is no such thing as mental causation of *actions*. *See Mental Cause Fallacy*, *supra* note 4, at 191, 218-21, 224-43. Stated bluntly, there is no need to explain the origin or content of thoughts in order to understand the *physiological* origins of acts.

53. *See supra* notes 40 and 46.

54. Compare Wittgenstein's admonition: "don't think, but look!" LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS § 66 (E.M. Anscombe. P.M.S. Hacker & Joachim Schulte, trans. 4th ed. 1953: 2009).

55. Contrary to Stephen Morse, statements like the ones in the text do not "["assume[] that all punishment is unjustifiably harsh because no one deserves any punishment at all." Stephen J. Morse, *Internal and External Challenges to Culpability*, 53 ARIZ. ST. L.J. 617, 648 (2021). What is assumed is the very different idea that "all punishment is unjustifiably harsh if no one deserves any punishment at all." *Id.* It assumed, in other words, that those who advocate the infliction of hardship and deprivation on their fellow human beings have the burden of persuasion on the question of whether offenders *qua* offenders deserve it. *See* JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION Chap XIII, sect. II clxvi (1780) ("all punishment is in itself evil").

IV. NEURAL DETERMINISM

Determinism is the idea that every physical movement that ever occurs is determined by prior physical states and the laws of nature.[56] Classic universal determinism is chiefly known as the "opposite" of so-called free will. To be clear, the findings of neuroscience do not prove the truth of classic universal determinism, nor do they directly disprove free will.[57] What the findings of neuroscience do supply, however, is overwhelming empirical evidence of neural determinism (or neurodeterminism). That is, neuroscience research shows it is practically certain, as a factual matter, that coordinated muscle contractions (and, hence, behavior) are proximally caused only by neuronal impulses, and that neuronal impulses are, in turn, proximally caused only by other neuronal impulses and, to a lesser extent, other physical factors acting on neurons (e.g., forces impacting on the sensory receptors and hormonal neuromodulators)—all of which activity is mechanismically determined in accordance with physical laws.[58] There is no evidence of any non-neuronal proximal cause for coordinated muscle contractions, bodily movements or human behavior, nor is there even a hint of one. Nor are there "gaps" in the neuroscience description of behavior-production that mental-state causes could fill.[59] Indeed, in the electro-

56. Carl Hoefer, *Causal Determinism*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2016), https://plato.stanford.edu/entries/determinism-causal/#ChaDet [https://perma.cc/VHC3-DQUY]. While I personally doubt the facticity of universal determinism, no position need be taken on that issue here since it is irrelevant to the question of mental causation, which could exist whether or not universal determinism is true.

57. That is to say, no claim is made here that neuroscience rules out the conjecture that there is some kind of interoception by which the brain and central nervous system are able to detect the qualia (contents) of self-generated mental states and prompt actions in accordance with them. I hasten to add, however, that neuroscience had turned up no evidence that such interoception occurs or that there are structures in which it might occur. *Cf. supra* text accompanying notes 39-40 and Andrea Lavazza, *Why Cognitive Sciences Do Not Prove That Free Will Is an Epiphenomenon*, FRONTIERS PSYCHOL. (Feb. 26, 2019), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6399109/ [https://perma.cc/F4BW-PD3S]. ("[O]ne can reasonably conclude that the data available are not sufficient to deny that we are endowed with free will in the form of conscious control that makes us morally responsible for what we do.")

58. Neural determinism, as defined in the text, does not (obviously) depend on the truth of classic universal determinism. What is more, unlike universal determinism, neural determinism does not make the large and difficult-to-verify claim that *everything* that happens is determined by what went before and natural laws. On the contrary, neural determinism makes only the much more parsimonious and essentially empirical claim that coordinated muscular contractions (behavior) are produced and determined solely by ordinary physiological processes.

59. *See* further discussion at *Mental Cause Fallacy*, *supra* note 4, at 238-43. *Cf.* Carroll, *supra* note 27 (observing that "[w]ithout dramatically upending our understanding of quantum field theory, there is no room for any new influences that could bear on the problem of consciousness"). Note that saying neuroscience explains behavior does not mean it can make exact predictions of what a brain (person) will do in concrete situations. The reason neuroscience cannot make accurate concrete predictions is that every human bodily movement is produced by an immensely complex collection of multifactorial causes and due to their very multiplicity, it is not possible to measure and weigh all the relevant factors. *See* Sapolsky, *supra* note 29, at 598-605. This impossibility does not mean, however, there is any experimental doubt about the neuroscience conclusion that the causative factors of behavior are exclusively physical. One may draw an analogy to the weather or stock prices, also multifactorial events: Nobody thinks the inability to measure and weight the multiple factors affecting the weather or stock price movements is evidence of non-physical forces at work. Neither should one suspect the

chemical chains of neuronal events that are the proximal causes of bodily movements, there is not even a place at which extraneous, nonphysical causes like intentions, reasons or a "free will" conceivably could intervene. There is no mechanism by which such nonphysical causes even *could* deflect the electrochemical chains of neuronal events from their nomothetically prescribed courses.[60] As a matter of empirical evidence, it is extremely unlikely that free will exists as a matter of *fact*.

None of this is to say that the neuroscience description of behavior production is complete. Though a great deal is now understood about neural network mechanics, no one has yet mapped the brain's neural networks in fine detail, and the architectures of their "algorithms" are far from reverse engineered. But even though neuroscience has much to learn, the operative physiological mechanisms that produce bodily movement, especially at the level of the neuron, are well understood and documented.[61] We do not know all the neuronal pathways, but we know what all the pathways consist of and that only *physical* causal events have ever been seen to trigger neurons into action. This evidence suffices to establish neural determinism to a near certainty.[62]

By contrast, there is no evidence whatever that any person has ever performed a bodily movement otherwise than when and as the person's nervous system directed the muscles to contract.[63] There is not only no evidence but not even a hypothesis as to how a person's "will" could organize and coordinate the millions of synaptic discharges that are physiologically requisite for any but the simplest spasmatic movement. In sum, the evidence leaves no doubt that the electro-chemical functioning and interaction of individual nerve cells is deterministic and that their operation in concert mechanismically determines the motor outputs of the brain's neural networks. This is neural determinism.[64]

---

preternatural in the case of behavior.

In short, the case for neural determinism of behavior is based on the empirically documented fact that coordinated muscle contractions are proximately caused and determined *only* by neuronal impulses and (except at sensory receptors) neuronal impulses are brought about only by other neuronal activity and other physical factors acting on neurons. Neural determinism parsimoniously rejects the postulation of any additional causative factors (such as mental states) for which there is neither evidence nor need.

60. *See* ARTHUR EDDINGTON, THE NATURE OF THE PHYSICAL WORLD 299-311 (Cambridge Univ. Press 1948), http://henry.pha.jhu.edu/Eddington.2008.pdf [https://perma.cc/99SH-S382], which I have discussed at greater length in *Mental Causation Fallacy*, *supra* note 4, at 219-21.

61. *See supra* notes 42 and 45.

62. *Id.*

63. For example, persons whose nervous system has been physically damaged or disabled cannot, by the force of mere intentions or "will," move the limbs that the damaged nerves would have served.

64. To be sure, the findings of neuroscience do not explicitly disprove the *possibility* of mental causation or free will. *See* Lavazza, *supra* note 57 ("[O]ne can reasonably conclude that the data available are not sufficient to deny that we are endowed with free will in the form of conscious control that makes us morally responsible for what we do."). And this non-disprovable *possibility* is a fact that sometimes drives friends of free will to become more than a little cock-a-hoop, *e.g.*, George Ellis, *From*

Obviously, the mountain of evidence for neural determinism creates a problem for any who want to use people's deeds as a basis for assessing moral worth or deservedness. If the internal causes of bodily movements are all physiological, and therefore subject to physical laws, then everything a person does would depend entirely on physical factors and chains of forces that can only come from outside the person herself, like the fateful wind at the regatta.[65] The reason one of Steinhoff's swimmers was an altruistic doctor and the other was a genocidal dictator can be seen, on the neural determinist account, to be not because of anything either of them could have mentally altered. Rather, the difference in their life outcomes can be seen as entirely due to the fact that they were products of (and conduits for) two different sets of causal chains, tracing back to events entirely outside themselves. The same can be said of the difference between Boris and Amos.[66] To make moral distinctions based on such adventitious differences

---

Chaos to Free Will (Here's Why so Many Physicists are Wrong About Free Will), AEON (Jun 9, 2020), https://aeon.co/essays/heres-why-so-many-physicists-are-wrong-about-free-will [https://perma.cc/KM88-2645] (asserting that "thoughts and feelings reach 'down' to shape lower-level processes in the brain by [somehow] altering the constraints on ion and electron flows in a way that changes with time."), .. . But even though science has no evidence that directly *disproves* mental causation or free will (or, for that matter, psychokinesis, extrasensory perception, goblins, ghosts and many other such preternatural processes), it has found plenty of evidence for alternative explanations of these apparently supernatural effects. All that science's "failure to disprove" shows is the near impossibility of disproving the existence of alleged colorless, odorless, weightless and non-haptic incorporeal presences. Indeed, all things considered, stressing the failure of science to disprove mental causation is rather desperate response to the mountain of neuroscience evidence for the alternative *physical* explanations of putative mental causes—like insisting that nothing in automotive engineering disproves the possibility of little car-spirits that push down the pistons when the gasoline flashes. Of course, such car-spirits would be undetectable (no mass, no odor, no space occupancy, no energy emission, etc.), just like mental causation, but they still cannot be *dis*proved.

Though lack-of-disproof is the weakest possible argument that mental causation exists, most of the philosophical literature I have seen in support of mental causation argues only that mental causes are *possible* (not ruled out), not that they actually exist.

65. *See supra* text accompanying notes 21-23. As Daniel Dennett recently put it, succinctly but unsympathetically, "nothing is ever in anybody's control." *See* DANIEL C. DENNETT & GREGG D. CARUSO, JUST DESERTS: DEBATING FREE WILL 84 (2021). Note that the external factors and forces that act on the person to determine behavior can have their effects either immediately or in the future—the latter occurring primarily due to the subsequent behavioral effects of microstructural physical changes in the brain (memories) that result from information acquired via the senses. Outside forces can, of course, also have other structural impacts on the body (e.g., loss of a body part) which can affect future behavioral choices.

66. In an effort to salvage some vestige of "self-authorship" despite determinism, Alina Roskies has pointed out, no doubt correctly, that persons are able to do things in the present that will affect what they will be, intend, desire, prefer, etc. in the future. *See* Adina L. Roskies, *Don't Panic: Self-Authorship Without Obscure Metaphysics*, 26 PHIL. PERSP. 323, 331 (2012) ("Through deliberately thinking and acting in strategic ways [in the present] we can exert control, modulate and intervene in our future states, both physical and mental. It is this that allows us to shape ourselves in ways that make it the case that we are in some very real sense responsible for who we are."). *See also* in a similar vein DANIEL C. DENNETT, FREEDOM EVOLVES 301-02 (2003). But, as very clearly explained by Gregg Caruso, no person has this ability at any given time unless the ability is conferred, at that time or a previous time, by the action of chains of causation originating outside the person. DENNETT & CARUSO, *supra* note 65, at 75-81. In other words, even though persons can often do things that will change their future selves, the extent of their ability to do so—and, hence, their future behavior—is entirely due to chains of causation originating outside themselves. (Dennett's response fails. *Id.* at 68-75. Inasmuch as the model of self-improvement that it presents does not, in the end, ever deny that every factor contributing to a

in their physical histories is arbitrary and invidious.[67] By providing an amply substantiated alternative to mental causation, neural determinism eliminates the need for the mental-cause conjecture and casts a deep shadow on the traditional rationale for using past deeds to judge the moral worth or deserts of persons.

V. "EXPLANATION COMPATIBILISM"

In his recent book on free will, Christian List argues that mental causation's practical usefulness in *explaining* behavior is reason enough to regard it as "real."[68] And, indeed, there is no denying that in everyday life, as well as the behavioral sciences, a person's intentions, desires, beliefs, reasons, and other such mental states are "explanations" that we all resort to when trying to make sense of what we and others do. Every child knows how socially important the difference is between doing something "on purpose" and doing it by accident. And if someone asks why Carla left the party early, it would be absurd to respond by talking about her neurons and synaptic strengths.

Because mental-state explanations are indispensable to a practical understanding of human behavior, List argues, it's at least as appropriate to attribute a person's behavior to her mental states as to her physiology[69]—even though he agrees that physiology underlies it all.[70] According to List,

---

person's formation is due to exogenous chains of causal forces.)

67. Again, I am not saying there is no moral basis for preferring Schweitzer over Hitler or Amos over Boris. Most of us probably feel intuitively, and strongly, that there is. I am only saying that that the basis would be better sought in other moral principles that "permit all sorts of differentiation." *See supra* text accompanying notes 17-20.

68. *See, e.g.*, CHRISTIAN LIST, WHY FREE WILL IS REAL 118, 74-77, (2019) (hereinafter "*Free Will is Real*") (stating that human action should be casually attributed to mental states if doing so is "explanatorily useful" or a "practical" guide for our actions). Others have made similar arguments. *See, e.g.*, Susan Haack, *Brave New World: On Nature, Culture and the Limits of Reductionism*, *in* EXPLAINING THE MIND (Bartosz Brosek, Lukasz Kwiatek & Jerzy Stelmach eds. 2018) ("It's all physical, all right, but it isn't all physics"); and *see also* Luke William Hunt, *Does Criminal Responsibility Rest Upon a False Supposition? No*, 13 WASH. UNIV. JURIS. REV. 65, 72-73, 77, 83 (2020) ("psychology and the physical sciences are fundamentally different modes of inquiry," and "nothing precludes psychology from defining its own mental ontology in the same way that physical science defines its ontology," and so both can have legitimacy as explanations "informing the laws conception of criminal responsibility"). *Cf.* Andreas Kuersten & John D. Medaglia, *Neuroscience and the Model Penal Code's Mens Rea Categories*, 71 DUKE L.J. ONLINE 53, (forthcoming) ("the reality of mental states is fundamentally a psychological-behavioral matter"). *Id.* at 81.

69. *Free Will is Real*, *supra* note 68, at, *e.g.*, 9, 50, 58, 63. *See also* Kuersten & Medaglia, *supra* note 68, in which the authors point out that one must use psychological-behavioral studies to establish that a particular kind of mental state exists and that a particular kind of neuronal activity is a marker of it *before* one can use the neuronal activity as a marker of the mental state. Thus, any neuronal explanation of a mental state is dependent, at least indirectly, on prior psychological-behavioral studies of mental states. This well-founded insight does not, however, affect the core claim of this article, namely, that every bodily movement (and, hence, all behavior) is nomothetically determined by neuronal activity (proximally) and chains of causal forces from outside the person (distally).

70. *See Free Will is Real*, *supra* note 68, at 58-64. *See generally id.* at 124-39 (asserting it would be difficult or impossible as a practical matter to adequately explain why people do what they do without

"the crucial question we must ask in relation to mental causation is this: Are human actions best explained by taking people's intentions to be their causes, or are they best explained by … something else, such as certain nonintentional physical states of the underlying brains and bodies?"[71] For List, the answer is clear: "[M]icrophysical brain states are," he says, "too fine grained to serve as the difference-making causes of human actions . . . [but] intentions do make a difference."[72] If "our best theories in the human and behavioral sciences are committed to the view that there is such a phenomenon as intentional agency[73] . . . we have every reason to take that phenomenon at face value"[74] and regard it as "real."[75] So even while List does not deny that, at a more fundamental level, human behavior is caused by the activity of neurons, he claims that mental causation and neuronal explanations are compatible. List's position—that mental causation is "real" even while accepting that neuronal causation is also real—may be referred to as "explanation compatibilism."[76]

There are, however, problems with List's reasoning. First, while List is entitled to use the word "real" any way he pleases, it is hard to shake the

---

taking into account their mental states, such as their intentions, beliefs and desires).

71. *Free Will is Real*, *supra* note 68, at 118 (emphasis added). *See also id.* at 113.

72. *Free Will is Real*, *supra* note 68, at 137. Note that, when writing of mental causation, List has chosen a conception of causation, called "difference-making causation," which contents itself (following Hume) to treat correlation as causation. *See infra* and *Free Will is Real*, *supra* note 68, at 132. In fact, "difference-making" causation is a misnomer because its definition does not require that causes actually "make" any difference at all, at least not in the usual sense of "make." That is, the definition of difference-making causation (according to List) does not require that the causes of an event include any mechanism that actually *does* anything to produce, generate or otherwise bring about their putative effects. A so-called "difference-making" cause need only be *correlated* with the effects. Specifically, List says, an event C should be considered the cause of an event *E* only:

1. If *C* were to occur, then *E* would occur.
2. If *C* were not to occur, then *E* would not occur.

Id. So when List says that mental causes are "real" causes of behavior, what he actually means is only that mental states are *really correlated* with behavior—not that they necessarily have anything to do with producing it. *See also infra* note 88.

73. *Free Will is Real*, *supra* note 68, at 74-77. *See also id.* at 117-18.

74. *Free Will is Real*, *supra* note 68, at 74.

75. *Free Will is Real*, *supra* note 68, at 8 ("a phenomenon qualifies as *real* if recognizing its existence is explanatorily indispensable: we would fail to give an adequate explanation of the relevant domain without recognizing the phenomenon in question"). I do not take issue with List's using the methodology of "inference to best explanation" to decide among competing inferences and, indeed, I have used it myself. *See Mental Causation Fallacy*, *supra* note 4, at 238-43. It has been my assumption, however, that there can be only one "best" explanation of any given set of data, and that it is the one "that best explains the totality of the relevant evidence and data [and is therefore] inferred to be the one that comes closest to the truth. *Id.* at 238-39. List is not entirely clear what his criterion of "best" is, but it seems to have more to do with convenience of explanation for certain purposes rather than with explaining the *totality* of the relevant evidence and data.

76. Broadly speaking, "compatibilism" is the view (which has many variants) that free will or, at least, moral responsibility is compatible with determinism. *See* Michael McKenna, *Compatibilism*, *in* STAN. ENCYCLOPEDIA PHIL. (2019). What all of the compatibilisms have in common is that they attempt to establish, on various theories, that a person can be deemed morally responsible for movements of her body even though the movements were pre-ordained, like a falling rock is pre-ordained to keep falling or orbiting planets to keep orbiting. "Explanation compatibilism" refers to the idea that mental causation (and, hence, the free will and moral responsibility that depend on it) are compatible with determinism, including neurodeterminism.

feeling that mental causation is not real in the same way that, for example, the behavior of electrons and exoplanets is real.[77] Remember that the question at hand is whether, as a matter of *fact*, some people have less moral worth than others and deserve to be treated accordingly. In defending the harsh treatment of human beings in punishment, we presumably want a justification that is rooted in material fact, not just any curve-fitting "reality" that happens to come in handy—even if the curve-fitting model seems to be consistent with certain kinds of data.[78] That is to say, we want an explanation of human conduct that is not just coherent internally but is also in harmony with our larger understanding of how events occur in the Universe as a physical system—a feature that non-materialist mental-state explanations decidedly lack. Such explanations may be fine as useful fictions for the behavioral sciences and everyday life, but as justifications for inflicting hardship and deprivations on human beings, fictions will not do.[79]

---

77. My own inclination would be to say that mental causation is "real" only in the sense that property rights, corporations or privity of contract are real. Indeed, as a legally relevant conception, it seems that mental causation is real in *precisely* the way that property rights, etc. are real.

More broadly, there are many things we say exist, and that appear able to affect the physical world, but that have no mind-independent existence or causal powers (unlike, say, electrons or exoplanets), for example, a dollar (as distinguished from the paper that sometimes represents it), a corporation, property rights (or ownership), other legal rights and relationships, legislative intent, and a person's intentions. *See* Annemarie Kalis, *No Intentions in the Brain: A Wittgensteinian Perspective on the Science of Intention*, FRONT. PSYCHOL. 5 (Apr. 26, 2019); A.B. Didikin, *Free Will, Action and Responsibility: Philosophical and Legal Analysis*, 48 TOMSK ST. U. J. OF PHIL., SOC. AND POL. SCI. 186, 191 (2019). Perhaps the best way to think of these kinds of "realities" is as "linguistic phenomena that construct a … reality" based on "pattern recognition" involving "patterns [of phenomena] that are extended in space and time." As an example, the statement "Charles intends to go to the party" means "that a certain pattern exists in the world; but this pattern is not itself the intention." Kalis, *supra* at 5.

78. To be sure, explanations based purely on "curve-fitting" and other such correlations can provide useful and coherent accounts of observed data even though their conceptual structures do not reflect the actual structure of reality. *See supra* text accompanying notes 77-79. For example, the travel of light waves through space used to be explained by a luminiferous aether, and the combustibility of wood was explained by phlogiston. Like mental causation, both the aether and phlogiston were colorless, odorless and weightless incorporeal presences that worked well to provide coherent explanations that fit the data of the time. However, again like mental causation, both the aether and phlogiston have been superseded by simpler explanations, ones that do not require ad hoc stipulations of unsubstantiated existences. Neither aether nor phlogiston is needed for a coherent explanation of the observed data and correlations, and there is no evidence that either one is part of "the ontological furniture of the natural world" (to borrow the wording of Charles Taylor, quoted in Gorski, *infra* note 88, at 33). The same can be said of mental causation.

79. But it may be objected, isn't physicalism likewise a fiction or, at least, ultimately non-confirmable? My answer is this: Though I do not doubt that physicalism (or metaphysical materialism) is probably true, the positions set forth in the text do not depend on it. To support those positions, all I need to insist is that the physical reality we infer to exist is, along with its attendant physical explanations, different in an important way from the putative realities and explanations that List infers to be "real." The difference is that the physical reality we infer to exist is, if you will, a "deep-mechanism" reality, one that is underlain with layers of explanatory mechanisms that reach down to the level of fundamental physics, and one that is broadly consistent and coherent across our entire generally shared understanding of the (admittedly inferred) physical Universe. By contrast, the reality proposed by List (causation by non-material mental states) is a shallow-mechanism reality, one that is inferred in order to serve a special-case, that has no underlying explanatory mechanisms and that is not consistent or coherent with

There is, however, an even bigger problem with "explanation compatibilism" in this context, one that is more serious than mere quibbles about the meaning of "real." It is the problem that emanates from the fact that any given set of data can have a potentially infinite number of compatible explanations,[80] and that not all explanations are equally good for all purposes.[81] For instance, it is perfectly fine for a driving instructor to explain that "pressing the brake pedal causes the car to stop." This explanation tells the student all she needs to know. But a mechanic trying to fix the car's brakes needs a different level of explanation—one that explains the system not just in terms of the pedal and stopping but that reveals the underlying mechanism. An explanation that is adequate for the student is not adequate for the mechanic while the explanation needed by the mechanic would be overkill for the student. Different purposes require different "levels" of causal explanation.

Similarly, mental-state explanations of human behavior, though useful for many purposes, simply lack the descriptive detail that is needed to demonstrate the connection we seek between a person's bad deeds and something other than just the person's physical body.[82] They lack such detail because, being inferred from correlations alone,[83] mental-state explanations elide and gloss over the nitty-gritty physiological facts of how behavior is produced. By "abstracting away" from the physical mechanisms of behavior production,[84] mental-state explanations systematically ignore

---

our generally shared understanding of the physical Universe. Explanations that are coherent with a deep-mechanism picture of reality are, among other things, far less likely to be ad hoc coincidences than those that are coherent only with a shallow-mechanism reality of the sort posited by List. Mental causation explanations, being devoid of any hint of undergirding mechanisms, could not be shallower: The only explanation anyone can give as to *why* it occurs is to say, "it just occurs." True, physicalism may also be subject to the objection that its mechanistic causal connections cannot be traced "all the way down" without conceding, at some point, that they too "just occur," *see* Laura Felline, *Mechanistic Causality and the Bottoming-out Problem, in* NEW DIRECTIONS IN LOGIC AND PHIL. OF SCI. (forthcoming). But it does not follow that physiological explanations of behavior causation are comparable to mental-causation explanations. The depth of physiological explanations and their broad consistency with other phenomena assures that they are not just ad hoc contrivances. *See also supra* note 49. *Cf.* Carroll, *supra* note 27 (observing that "[t]o start with the least-well-understood aspects of reality and draw sweeping conclusions about the best-understood aspects is arguably the tail wagging the dog").

80. *See* Lee McIntyre, *Who's Afraid of Supervenient Law? in* ESSAYS IN THE PHILOSOPHY OF CHEMISTRY (Eric Scerri and Grant Fisher, eds. 2016) (stating that "the descriptions and theories that we use to capture … ontology … are potentially infinite …. There may be one and only one reality but there are an infinite number of ways of describing it"). For example, "there are more than 700 different versions of the periodic table [in chemistry], but only one periodic law" viz. "when arranged according to their atomic number – after certain regular but varying intervals the chemical elements show an approximate repetition in their properties." *Id.*

81. *See* McIntyre, *supra* note 80 ("depending on which descriptive terms we use regularities may emerge given some ways of looking at the world that will elude us using others").

82. *See supra* text accompanying note 16-32 ("Using Past Deeds to Assess Moral Worth or Deserts"). The bad deeds have to connect to something other than just the person's physical body because the physical body is (presumably) subject to physical laws—so every move it makes is a conduit for the forces that act upon it—and persons are not morally responsible for events that are dictated by physical laws. *Id.*

83. *See supra* text accompanying notes 41-43.

84. *See* Elliott Sobor, *The Multiple Realizability Argument Against Reductionism*, 66

the chains of physical events and micro-events that trace back from a bodily movement through the person's body to the movement's predominantly sensory triggers. Indeed, List avers, one of the great virtues of mental-state explanations is that they *do* elide and gloss over the mountains of daunting and distracting physical details (of synapses, networks, neuronal chemistry, etc.) which, in ordinary daily interactions, are of no interest.[85]

But no matter how useful this simplifying virtue of mental-state explanations may be for everyday and social science purposes, it is precisely their vice when it comes to the question of ascribing moral responsibility, of ascertaining whether bad deeds have anything to do with a person's moral worth or deserts. For if mental causation is going to be able to show there is a connection or nexus between a person's bad deeds and her mental states, it cannot be defined as merely a matter of *correlation* with no hint of corroborating mechanism.[86] In other words, one cannot define causation in a way that entails no ontological connection between causes and effects (as List does[87]) and then, at the same time, think that causation shows a connection between the two.[88]

PHILOSOPHY OF SCIENCE 542 (1999). *See* Katrina Sifferd, *Non-Eliminative Reductionism*, *in* Bebhinn Donnelly-Lazarov, NEUROLAW AND RESPONSIBILITY FOR ACTION 71, 100-101 (2019) ("Higher-level descriptions may *"abstract away" from* the physical details that make for differences among the micro realizations that a given higher-level property possesses") (emphasis added).

85. *See Free Will is Real*, *supra* note 68, at, *e.g.*, 58-63, 69-74, 131. If someone asks, "Why did X do A?", the questioner does not want to hear about synapses and neural networks but, rather, about intentions, desires, reasons, and other mental states. Likewise, if one asks, "Why does my computer display everything underlined?", one does not want to hear about program-lines and memory locations, but about which keystrokes activate and deactivate the underlining feature.

86. As per Glennan: "A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon" STUART S. GLENNAN, THE NEW MECHANICAL PHILOSOPHY 17, (2017). *See also* Jon Williamson, *Mechanistic Theories of Causality Part II*, 6 PHIL. COMPASS 421 (2011); Carl Craver and James Tabery, *Mechanisms in Science*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY 3.2.3 (2015).

87. *See supra*, note 72 and *supra*, note 88.

88. To take an obvious example, mental causation would not serve to provide the nexus needed for moral responsibility if a person's intentions merely happen to *coincide* with the person's deeds, without actively making them occur, or if both the deeds and the intentions were both brought about by the same third factor. *Cf.* Harry Frankfurt, *Alternative Possibilities and Moral Responsibility*, 66 J. PHIL. 829 (1969).

The root of List's difficulty is, I think, his adherence to a "difference-making" conception of causation, which treats epistemic *criteria* of causation (i.e., correlations) as though they tell us something about the ontology (connecting mechanism) of causation. *See supra* note 72. But, as they say, "correlation is not causation" and correlation-based mental-causation "explanations" do not even purport to provide evidence or a hypothesis about the underlying *mechanism* that is usually demanded for claims of causation in the case of biological phenomena—which the bodily movements of living beings most emphatically are. *See generally* Philip Gorski, *Causal Mechanisms: Lessons from the Life Sciences*, *in* GENERATIVE MECHANISMS (Margaret Archer ed. 2015). List's difference-making conception of causation, in which the causation is evidenced by (and, perhaps, consists of) correlations alone (or, per Hume, "constant conjunction"), *see supra* note 72, does not supply the univocal evidence of the ontological connection that we need to connect a person to her deeds and establish that she is morally responsible for them. It is incoherent to think both that "causation" need not entail a connection between causes and effects and, at the same time, think that causation supplies the connection between mental states and deeds that is required for attributing responsibility.

Indeed, when mental causation is inferred from correlations alone, a mental-causation explanation of behavior is not strictly speaking an *explanation* at all but only a "black box," a placeholder in lieu of explanation that stands between a putative cause and its putative effect. Because no one knows what's in the black box, such mental-causation explanations provide no basis at all for concluding that bad deeds are factually connected or attributable to the mental states (*e.g.*, intentions) that seem to portend them. The correlations we observe between the deeds and the mental states could just as easily be due, for example, to the fact that both the deeds and the mental states are the results of some third factor, e.g., exogenous chains of causal forces that act on the person (mostly via the senses) and, by means of the person-as-conduit, bring about both. As a result, a mental causation "explanation" of behavior is fatally ambiguous on the very point for which is being invoked, to demonstrate a connection between a person's mental states and the things the person does. Without such a connection, the person's deeds are not suitable criteria for judging moral worth.

In sum, List may be justified in claiming that mental-state explanations of behavior are "real" and are compatible with physiological explanations, but it does not follow that List's mental-state explanations provide a basis for attributing agential responsibility to persons or assessing their moral worth. Because mental-state explanations are based on correlations alone, without a hint of a corroborating mechanism, they lack the descriptive detail that is needed to avoid being fatally ambiguous on the question at hand, namely, whether a person's deeds are factually connected to the person's self-generated intentions and other mental states. If the person's deeds and mental states are not so connected, then the deeds are not suitable criteria for judging moral worth or deserts. Accordingly, whatever one thinks of explanation-compatibilism for some purposes, it does nothing to undercut the conclusion that a person does what she does, not as an agential initiator or author of her acts, but as a conduit for exogenous causes originating elsewhere.[89] So even accepting the validity of explanation compatibilism, mental-state explanations still would not support the assumption (apparently made by Steinhoff and Kagan), that bad deeds affect or reveal the moral worth or deserts of the persons who do them.

---

89. Actually, this highlights a flaw appearing to affect just about all forms of compatibilism: If determinism is true, as compatibilism assumes, the persons always do what they do as a conduit for chains of exogenous causes from outside themselves and, unless mental states (e.g., reasons) have thaumaturgic powers to confer moral significance to motions and events, the behavior produced via mental causation would no more reflect the moral worth of the person who performs it than bodily movements produced by physical causes alone—such as the gust of wind at the regatta.

VI. THE STRANGE PERSISTENCE OF MENTAL CAUSATION BELIEFS

If mental causation is neither substantiated in fact nor coherent with the larger picture of how physical events happen in the Universe, how do we explain its persistence as a core assumption of criminal law? During the law's formative years, the idea of mental causation was, of course, nothing exceptional. In those days it was perfectly normal to assert all manner of spooky-spiritual nonphysical explanations for various events.[90] Mental causation fit right in. Today, however, nonphysical explanations seem bizarre, especially when there's a plausible physical alternative at hand. A thump in the night makes most people think of something falling over, or maybe a burglar, but not of spirits and ghosts. Given the materialist, evidence-oriented tenor of the times, it is hard to see mental causation as anything but an ontological outlier—a misty-murky nonmaterial explanation of a kind that is generally eschewed in modern mainstream scholarship. No serious scholar in any other secular context would countenance such a thing.

So how does mental causation continue to be so thickly woven into criminal law?[91] How does it remain virtually unquestioned as a reason for the law to treat millions of people (whom we piously declare to be "created equal") as though they are morally inferior and deserving of hardship? Undoubtedly, a big part of the answer has to do with the pivotal role that mental causation plays in the prevailing logic of fault and moral responsibility. Without mental causation, guilt would have to be based on physical causes alone. For many, that would make current punishment practices hard to justify. We do not hold machines or automatons morally responsible or make them suffer for their wrongs. Fault presupposes that "the origin of an action is in oneself [and] it is in one's own power to do it or not."[92] In other words, what makes responsibility possible and

90. As recently as 1765, the great common law commentator William Blackstone believed in witchcraft (or, at least, felt compelled to profess that he did). 4 WILLIAM BLACKSTONE, COMMENTARIES, Ch. 4. (1765).

91. *See generally* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 117–44 (6th ed. 2012); WAYNE R. LAFAVE, CRIMINAL LAW 252–88 (5th ed. 2010); Rollin M. Perkins, *Rationale of Mens Rea*, 52 HARV. L. REV. 905 (1939); 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. *See also* Stephen J. Morse, *The Inevitable Mind in the Age of Neuroscience* 34, *in* PHILOSOPHICAL FOUNDATIONS OF LAW AND NEUROSCIENCE (Patterson et al. eds., 2016); 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) ("Roughly speaking, the law implicitly adopts the folk-psychological model of the person, which explains behavior in terms of desires, beliefs and intentions.").

92. ARISTOTLE, NICOMACHEAN ETHICS 1110a 119 (H. Rackham, ed.). As act is not deserving of praise or blame "when its origin is from without, being of such a nature that the agent, who is really passive, contributes nothing to it…." *Id. See also* OLIVER WENDELL HOLMES, THE COMMON LAW 54 (2009 [1881]) ("it is felt to be impolitic and unjust to make a man answerable for harm, unless he might have chosen otherwise"). *See* James W. Moore, *What Is the Sense of Agency and Why Does it Matter?*, FRONTIERS IN PSYCHOLOGY 7, doi: 10.3389/fpsyg.2016.01272 (2016), *available at*

punishment morally palatable is the belief that the choice and intention to do wrong are self-generated within the wrongdoer as mental states whose occurrence and content are not dictated by physical laws. By maintaining that offenders make their own behavioral choices, the law and its functionaries can disclaim moral responsibility for the hardship and deprivation that punishment inflicts. Those who incur suffering at the hands of the law bring it on themselves.[93]

By contrast, neural determinism (the only real alternative to mental causation) is squarely at odds with the possibility of self-generated behavioral choices. Even though the neuroscience explanation of behavior is amply documented by numerous studies,[94] it leads to conclusions that many find disagreeable, an affront their moral sensibilities and to the prevailing punitive ideology.[95] Specifically, it leads to the conclusion that persons are mere conduits for forces that act upon them from outside, all in accordance with physical laws.[96] And, if that is so, then the origin of bad actions is *not* within the persons who do them, and it is not within their "own power to do it or not."[97] Obviously, there are many who simply do not *want* to believe conclusions like these. They do not want to believe the neuroscience explanation of human behavior, for it implies[98] that their punitive impulses (and current punishment practices) are morally wrong.

In this context, the fact that mental states are observably "anomalous"[99] makes mental causation supremely attractive as an explanation of criminal behavior. Anomalousness, you will recall, means mental states are not explainable or predictable by strict deterministic laws.[100] By their very

---

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5002400/ [https://perma.cc/PT3M-FGN9] ("for most people it only makes sense to hold someone responsible for their actions if they are freely in control of them").

93. Just to be clear, in floating these "can do otherwise" criteria for attributing responsibility and inflicting punishment I do not mean to be endorsing them as well. For a "basic-moral-equality" view of conditions that must be met to justify punishment and responsibility, *see infra* text accompanying notes 116-20 and *Mental Cause Fallacy*, *supra* note 4, at 246-53.

94. *See* Davidson, *supra* note 36.

95. *See* discussion in Andrea Lavazza, *Neurolaw and Punishment: A Naturalistic and Humanitarian View, and its Overlooked Perils*, 37 TEORIA 81, 93 (2017) ("The evolutionary processes of the species, driven by selection and adaptation, have endowed us with very strong intuitions – generally retributive – that cause people to be ready to bear a personal cost, with no other gain than the restoration of a sense of justice, to punish offenders who deserve it").

96. *See supra* text accompanying note 25-31.

97. *Cf. supra* text accompanying note 92.

98. Because neuroscience studies only physical facts, not moral precepts, it does not make normative judgments about right and wrong. But when normative judgments are predicated on physical facts, as they always are, it is important to get those facts right. For when normative judgments are based on erroneous facts, they can be just as wrong as ones based on erroneous moral standards. The neuroscience explanation of human behavior implies that common punitive impulses and punishment practices are morally wrong by showing they are predicated on erroneous assumptions of fact (viz. the facticity of mental causation).

99. *See supra* text accompanying note 36. Note that if mental states are truly anomalous as well as causally efficacious, then explanation compatibilism cannot be true. Behavior cannot be both dictated in accordance with physical laws (determinism) and not so dictated (autonomous).

100. See *supra* text accompanying note 36.

existence, therefore, anomalous mental states seem to stand as proof that neuroscience's deterministic explanation of behavior cannot be complete. In fact, anomalous mental states look to be an outright *exception* to determinism, incontrovertible evidence that determinism cannot be *entirely* true. What is more, if mental states are anomalous, they would not need to have sufficient prior causes for their occurrence or content. This would mean there is nothing that prevents mental states from being self-generated within the person herself—out of nowhere, so to speak. And so, it seems to follow, if a person's self-generated intentions, desires, and other anomalous mental states, can insert themselves into the body's neuronal chains of causation (as people fallaciously infer[101] that they can), then the neurodeterministic "conduit" model of behavior can be sidestepped. We would be able to say that, at least sometimes, a person's conduct originates agentially (albeit somewhat mysteriously) within the person herself, as anomalous mental states. That, in turn, would mean a person's bad deeds could be seen as her own and, as such, could serve as suitable criteria for judging the person's moral worth or deserts.

In short, the fact that the occurrence and content of mental states are observably anomalous, untethered to physical law, coupled with a belief that mental states are causally efficacious, surely has much to do with the reluctance to let go of mental-causation beliefs, despite their dubious factual basis. Simply put, mental causation beliefs are so persistent in law because they help the law duck the neuroscience threat to offender responsibility and traditional punition.

There are, however, at least two major problems with this reasoning. One of them, already discussed, is the absence of any evidence or valid inference that there *is* such a thing as mental causation.[102] Mental states cannot have moral efficacy if they do not have causal efficacy. A person's intentions and other mental states will not justify judging the person by her deeds if the mental states have nothing to do with what the person does.

The second problem with this reasoning is its assumption that the anomalousness of mental states means they also are *autonomous*, *viz.* not determined in accordance with physical laws. It assumes, in other words, that mental states do not depend for their occurrence or content on deterministic chains of causal forces (such as neuronal activity in the physical domain) but that they are potentially self-generated out of nowhere, so to speak. The reasoning needs to make this assumption that mental states are autonomous because, to the extent their occurrence and content are *not* autonomous (but are determined, for example, by a physical neuronal substrate), a person's behavior would not be her own. It would be, rather,

101. *See supra* text accompanying note 41-41.
102. *See supra* text accompanying notes 38-43.

entirely determined by the chains of physical forces that produced it. If causally efficacious mental states are not autonomous, they are merely part of the conduit that conducts the causal chains that produce human behavior.

But the assumption that because mental states are anomalous they must *a fortiori* also be autonomous is false. Anomalous and autonomous are not the same. Indeed, nature abounds with events that are observably *anomalous* (not explainable or predictable by strict deterministic laws) but are nonetheless produced by deterministic physical processes that *do* conform to strict deterministic laws—for example, the timing and intensity of earthquakes, the shapes of mountains, the distribution of oceanic islands, next Thursday's weather, and just about everything that is studied by "chaos theory." There is nothing in principle that prevents a computational (mechanismic) information-processing system, though itself subject to physical laws, from producing representations of fanciful alternative "realities" that contain events and scenarios could never really exist. Interactive video games are a familiar example. These games often generate and portray progressions of events that could never occur under physical laws even though those portrayals are produced by computers whose workings, like those of any physical system, are wholly deterministic in accordance with those laws.[103]

Similarly, even though anomalous mental states can wander non-logically through our conscious mental space and create intrinsically inconsistent and contextually incoherent depictions of a "mental world," this anomalousness does not mean they do not depend on underlying nomothetic physical (neuronal) processes for their occurrence and content. In other words, the anomalous nature of mental states provides no "proof" whatsoever that determinism has an exception, and no proof that a person's mental states arise inside the person autonomously self-generated out of nothing. Most importantly, the anomalousness of mental states is of no help at all in establishing that a person is anything but a conduit when she chooses what to do, a conduit for exogenous forces that determine her every move. And, as we have seen, the deeds a person does as a conduit for exogenous forces are not a suitable basis for assessing her worth or deserts.

Still, one might object, nothing in neuroscience proves that mental states actually *do* depend on deterministic physical processes (such as the brain) for their occurrence and content. And certainly, nothing proves they *invariably* do. To the extent that mental states are *not* dependent on the brain, they are almost surely autonomous, which means that the behavior they cause can be properly regarded as the doer's own and, as such, a suitable criterion for judging moral worth and desert.

These are valid points. While most neuroscientists and other students of

103. Indeed, reasoning along these lines seems to be at the root of most compatibilism: the idea that observably anomalous causative mental states can somehow emerge out of a Universe of physical processes whose deterministic character is not in dispute.

the mind seem to assume that the mind is produced by the brain, the evidence for this position is at best correlational, and the mechanism for creating mental states has not yet been found. Thus, the conjecture that mental states may be autonomous, at least sometimes, cannot be ruled out. Embracing the conjecture would, however, create some wrinkles that need attention.

For one thing, most of the usual arguments for mental-state causal efficacy would no longer apply. For example, if mental states are truly autonomous of physical brain states, it would no longer make much sense to explain their causal efficacy by saying they are "supervenient" on[104] or properties of physical brain states. Nor would it make sense to suggest, as John Searle has, that mental states have causal powers because mental states and brain states are just two levels of description of the same thing.[105] Indeed, all of the so-called "identity theories" (that mental states are identical to brain states) would go out the window.[106] In other words, if autonomous mental states do have causal efficacy, it would have to be in their own right, not because they are aspects, properties or supervenients of the physical brain.[107]

Other implications of autonomous mental states are perhaps even more unsettling: if mental states do not depend on the brain for their occurrence or content,[108] why would they need the existence of a living brain at all? Or, for that matter, a living body? Do autonomous mental states imply the real possibility of disembodied spirits?

As one can see, pursuing the idea of autonomous, self-generated mental states quickly gets one way beyond the evidence. The philosopher Michael S. Moore derides the whole idea that mental states can be autonomous yet efficacious as ghosts that "can throw real rocks but somehow cannot be hit

---

104. Roughly, mental states are considered "supervenient" on brain states if they are dependent on the brain states and there could be no variation in the mental states without a corresponding variation in the underlying brain states, *See* Brian McLaughin & Karen Bennett, *Supervenience*, STAN. ENCYC. PHIL. (2018), https://plato.stanford.edu/entries/supervenience [https://perma.cc/Y8F7-C6L6]. My own view is that mental states almost certainly are supervenient on brain states and therefore not autonomous (as I think is the dominant view among specialists concerned about these issues). This cannot, however, be taken as a confirmed fact.

105. JOHN R. SEARLE, MIND: A BRIEF INTRODUCTION 210 (2004).

106. As perhaps they should. *See* JAEGWON KIM, PHYSICALISM, OR SOMETHING NEAR ENOUGH 121-48 (2005); *See* JAEGWON KIM, PHILOSOPHY OF THE MIND 62-71 (1998). As an example of an appeal to identity theory, *see* Michael S. Moore, *Libet's Challenge(s) to Responsible Agency*, *in* CONSCIOUS WILL AND RESPONSIBILITY, 207, 225, 227 (Walter Sinnott-Armstrong & Lynn Nadel eds., 2011) ("willings, being identical to some swatch of the chain of brain events that cause voluntary bodily movement, will be the initiators of action, just as the folk psychology and morality supposes").

107. This is essentially the hypothesis known as "substance dualism," which was kind of mind-body dualism famously expostulated by Descartes and is now widely considered discredited. *See* Justin Skirry, *René Descartes: The Mind-Body Distinction*, INTERNET ENCYCL. PHIL., https://www.iep.utm.edu/descmind [https://perma.cc/M7U7-XFES] (last visited Feb. 3, 2022).

108. *See* Humbach, *supra* note 33, at 5-6, 13-18 (discussing reasons to think that mental states are *not* autonomous of the physical).

by them."[109] For a hard-headed materialist steeped in today's science, it is probably just about as hard to see how mental states could occur out of nowhere as to see how they could move the human body. In any case, the singular nature of autonomous-but-efficacious mental states looks suspiciously ad hoc—a kind of spooky-spectral numinous nonsense that no modern thinker, in any other context, would accept. As such, they are a slender reed on which to rest a theory of punitive "justice."[110]

To summarize, one reason mental-causation beliefs have such continuing appeal is that mental states are observably anomalous, which not only suggests that determinism cannot be *entirely* true but also seems to offer a plausible escape from the neurodeterminism threat to current responsibility assumptions and punishment practices. But the escape is illusive. Anomalous does not mean or necessarily imply autonomous and, for intentions or other mental states to be truly the person's own, they would have to be not just anomalous but autonomous. To the extent that causative mental states are not autonomous in their occurrence and content but are instead "determined" as per physical law, they would be no more morally relevant than the gust at the regatta.[111] It would make no more sense to blame a person for her bad deeds than to blame your computer because you do not like the news you see on the internet.[112]

VII. WHAT ABOUT DANGEROUS PEOPLE?

To recognize and respect the basic moral equality of all would mean a major rethinking of criminal justice practices, but it would not change the fact that there are dangerous people—individuals who pose unreasonable risks of harm to others. Obviously, such persons cannot be simply left to do their worst. Basic moral equality not only allows but probably requires that society protect itself and its members by restraining those who would encroach on the equal rights of others.[113] After all, if moral equality means anything, it means that no one is morally privileged to do things that harm

109. Moore, *supra* note 43, at 57.

110. Mark Balaguer argued that anomalous ("free," undetermined) mental states are able to cause behavioral choices despite the "determinism" and "epiphenomenalism" arguments to the contrary. In the end, however, Balaguer seems to conclude there is neither much reason to believe or to disbelieve that cognitive decision making can be both anomalous ("free") and physically efficacious (non-epiphenomenal). *See* Mark Balaguer, *Free Will, Determinism, and Epiphenomenalism*, 9 FRONTIERS PSYCHOLOGY 2623 (2019.).

111. *See supra* text accompanying notes 21-31.

112. I.e., to the extent the occurrence and content of causative mental states are not autonomous but are for example determined by their physical (neuronal) substrate, bad deeds cannot be thought to be the person's own but are, rather, determined by the chains of physical forces that produced them.

113. The principle is roughly this: No one has a moral right to violate the equal interests of others, and it is therefore not morally wrong to restrain another person from doing so. *See* SCHOPENHAUER, THE WORLD AS WILL AND IDEA, *supra* note 2, at 342 ("[W]hat is done simply in order not to suffer any wrong is not wrongdoing"); *see generally id.* at §62, especially 339–42. *See also* FREDERIC BASTIAT, THE LAW 2 (Dean Russell trans., 2011) ((1850) ("[T]he principal of collective right … is based on individual right."); *Mental Cause Fallacy*, *supra* note 4, at 249-50.

or threaten others' equal rights to not be encroached upon or used as means.[114] There is no privilege to pose a danger to others even if the danger is due to causes from outside oneself. Acts can be morally wrong even if they are neurodetermined.[115]

In some cases, the only way to protect the equal moral rights of all is to incapacitate and perhaps even confine those who pose unreasonable risks of harm to others. Modern criminal justice, especially in the United States, makes abundant use of confinement as well as other incapacitation, but current practices have vast room for improvement when it comes to respect for the basic moral right of everyone, including offenders, to equal concern and decent treatment. We must protect ourselves from dangerous people, but we do not have to torment them.

What basic moral equality means is that even the interests of offenders are entitled to "equal concern and respect."[116] It means that, though some must be confined or coerced for the protection of others, even they have a basic and equal right to not be treated as objects or as means. The overriding goal of corrections should be, in other words, to respect and preserve the personhood, interests, and equal moral worth of every individual except to the extent that interference is *inseparable* from measures to prevent encroachments, in excess of one's right, on the equal rights of others. Though the loss of free ubiety (imprisonment) may sometimes be a regrettable social necessity, the need to confine does not justify any greater hardship than confinement makes inevitable.

Treating offenders with "equal concern and respect" means, among other things, that the living conditions of prisoners should be comfortable and dignified, with no greater impingement on their pursuit of their own human needs, desires and associations than is inseparable from the needs of public safety.[117] Everything in the correctional experience should be geared to promote successful re-entry into society because, among other things, re-entry is usually inevitable and unsuccessful re-entries are serious threats to the public. And, of course, confinement should be avoided entirely whenever there are other measures at hand that will be just as effective in preventing crime.[118]

This mild treatment of offenders may sound bizarre to ears accustomed to today's harsh views of offenders and punishment. But it follows from

---

114. *See supra* note 2.

115. As further discussed *infra* text accompanying notes 115-35.

116. RONALD DWORKIN, SOVEREIGN VIRTUE: THE THEORY AND PRACTICE OF EQUALITY 406,411 (2000) ("equal concern and respect").

117. *See generally* my further discussion of punitive practices in *Mental Cause Fallacy*, *supra* note 4, at 250-52.

118. *See* Brian Galle, *The Economic Case for Rewards Over Imprisonment*, 96 IND. L.J. 471 (2021) (arguing that, when you do the math, it would save money and be more effective to largely replace today's punishment approach crime reduction with a less expensive "rewards" approach, concluding that "if cuts were made with any care at all, we could save billions without increasing crime").

recognition of basic moral equality and an understanding that bad deeds do not reflect the doer's moral worth or deserts. To torment human beings beyond the needs of public safety, though a routine function of government today, is an irrational expression of hatefulness and an abnegation of basic moral equality.[119]

The current Covid-19 pandemic provides a perspective. As of this writing, approximately 4,000,000 people in the United States have been *afflicted* with the disease (roughly, those needing hospitalization)[120] while many millions more, perhaps tens of millions, have been asymptomatic, pre-symptomatic or pauci-symptomatic carriers.[121] It is reasonable to assume that many of these latter millions have, due to causes from outside themselves, transmitted a deadly and debilitating disease to others—one that has had a fatality rate of as much as a 5-6% among those actually afflicted (i.e., who are not mere carriers).[122] Like people who commit crimes, these millions of carriers are literally conduits for deadly and debilitating forces that come from outside themselves. While most would probably agree that infectious carriers of deadly diseases should be subject to coercive measures to protect others, even including quarantine, few would say that those with infections have lower moral worth or deserts, much less that they should be tormented along the lines that are routine for other hapless conduits of harmful forces.[123]

---

119. In response to the obvious question of "who pays for all this?," there is no reason to conclude, without doing the math, that the cost will be disproportionate to current outlays or that it will be unmanageable. *See* discussion of the point by GREGG D. CARUSO, REJECTING RETRIBUTIVISM: FREE WILL, PUNISHMENT, AND CRIMINAL JUSTICE Ch.9 (2021). What is more, consistent with basic moral equality of concern and decent treatment, there is no reason to prevent persons in custody from engaging in efforts at self-improvement and earning for themselves—to pay for the comforts (above a basic level) that they choose to enjoy. In addition, removing retributive bases for punishment would likely result in shorter terms of incarceration, another source of savings.

120. CDC, *Coronavirus Disease 2019 (COVID-19) COVID Data Tracker*, CDC (Apr. 30Aug. 1, 2020), available at https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html [https://perma.cc/327M-NKQF]; *see also* NCIRD and Division of Viral Diseases, *Covid Data Tracker Weekly Review*, CDC (undated), https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html [https://perma.cc/9S9V-P6FD].

121. *See, e.g.*, Gretchen Vogel, *Antibody surveys Suggesting Vast Undercount of Coronavirus Infections may be Unreliable*, SCIENCE (Apr. 21, 2020), https://www.sciencemag.org/news/2020/04/antibody-surveys-suggesting-vast-undercount-coronavirus-infections-may-be-unreliable [https://perma.cc/PV6E-LCHH].

122. *See* chart posted at, Hannah Ritchie et al., *Mortality Risk of COVID-19*, OUR WORLD IN DATA (2020), https://ourworldindata.org/mortality-risk-covid [https://perma.cc/6MQ5-PA5H]). Fortunately, the mortality rate is now considerably lower. Id.

123. *See* CARUSO & DENNETT, *supra* note 65, at 127-35; Gregg D. Caruso, *Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model*, 32 S.W. PHIL. REV. 25, 28–31 (2016).

While I am not ready to advocate pharmacological "moral enhancement," *see* Ingmar Persson & Julian Savulescu, *The Evolution of Moral Progress and Biomedical Moral Enhancement*, 33 BIOETHICS 814 (2019), its growing technological plausibility will make it increasingly a factor to reckon with in future determinations concerning the dispositions of offenders. While ethical considerations are mentioned against it, Ignacio Macpherson, María Victoria Roqué et al., *Moral Enhancement, at the Peak of Pharmacology and at the Limit of Ethics*, 33 BIOETHICS 992 (2019), these seem to be premised on the idea that harsh measures to reform, rehabilitate, deter, etc. are morally superior to gentler pharmacological ones.

In short, though we must deal with dangerous people, the justifiability of current criminal justice practices is, in the light of basic moral equality and the findings of neuroscience, highly questionable. But the problem is not that neuroscience shows (in the words of Stephen Morse) that "all punishment is unjustifiably harsh because no one deserves any punishment at all."[124] The reason current criminal justice practices are questionable is that, due to basic moral equality, no one has a special moral privilege to inflict hardship and deprivation on any other person. Thus, those who presume to inflict or claim a right to do so need to offer better evidence of justification than just their unsupported belief that minds cause bad deeds.[125]

VIII. IF NEURODETERMINISM IS TRUE, HOW CAN ACTIONS BE MORALLY "WRONG" (OR "RIGHT")?

A consideration that is frequently raised against determinism, including presumably neural determinism, is that it erases the distinctions between "right" and "wrong," at least in any moral sense. If the world runs like a clockwork and the drivers of human behavior all originate outside the person, how can any human action be sensibly praised as good or condemned as morally wrong? How, for example, can one morally criticize the way that the law treats offenders if that treatment is simply the way things are, determined and dictated by forces outside anyone's control? Doesn't the neural determination of behavior mean that every human action, like all other physical events, is morally neutral—no more right nor wrong than anything else that happens in the Universe? This is not the place for a full response to these questions but, for the sake of context, a few comments are in order.

First, the moral quality of a person's actions depends on factors that neuroscience simply does not address. Neuroscience is concerned with *physical* events, and nothing in its findings confirms or denies that human actions have moral valence, that they can be good or bad, right or wrong.[126] The factors on which rightness or wrongness depend are matters of philosophical debate among, for example, utilitarians, Kantians, virtue theorists, moral realists, contractarians, coherentists, various religious sectarians and others, and those factors need not be debated here.[127] What

124. Stephen J. Morse, *Internal and External Challenges to Culpability*, 53 ARIZ. ST. L.J. 617, 648 (2021).

125. *See* JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION Chap XIII, sect. II clxvi (1780) ("all punishment is in itself evil").

126. Neuroscience research can only discover what "is," and an "is" does not imply an "ought." DAVID HUME, TREATISE OF HUMAN NATURE 469 (1826).

127. With the partial exceptions of *infra* text accompanying notes131-37.

My own preference, consonant with basic moral equality, would be a form of Kantianism, *see supra* note 2, or "semi-Kantianism.", *See* Benbaji, *supra* note 2, at 473 ("no person is subordinate or superior to another person.").

suffices for present is to note that, even if persons are not the authors and initiators of their own acts, there is no reason that the acts themselves cannot be praised as right, condemned as wrong, and so on. The crucial point is this: it can be morally wrong for *Anna* to cause suffering to *Berle* even if it makes no sense to judge *Anna* as morally inferior for doing so (because, for example, she acted as a conduit for physical causal forces from outside herself). Nor does anything in neuroscience make it pointless to critique bad behavior. For even if human actions are fully neurodetermined, moral discourse (such as praise, condemnation, and other critique) could still be hugely significant as means to shape behavior. Moral discourse works to shape behavior for the same reason that deterrence works, namely because the things that happen to people (including reasoned arguments and critiques) can affect what people do. But that does not mean there is anything to connect the moral worth of a person's actions to the moral worth of the person herself: Actions can be right or wrong, or good or bad, but the person who does them has a basic moral entitlement to equal concern and respect solely by virtue of being human.

Second, even if persons are not the causal origins of their acts and do not mentally control what they do (except as a conduits), it does not follow that a person is essentially just a puppet, devoid of moral worth. Persons are conscious beings that have rich phenomenological lives—lives that can be filled with happiness or misery, hopes or despair, fears or comfort and all the rest, depending on how they and others behave. As such, human beings are, obviously, ontologically different from puppets or dolls. And, although this ontological difference does not logically *entail* that there is a moral difference as well (an "is" does not imply an "ought"[128]), it is likewise risky to assume that people and puppets are morally equivalent *despite* the sharp ontological difference between them. Given the overwhelming importance of the phenomenology of consciousness to our lives (it is, after all, the only thing anyone ever directly knows), it matters immensely to the quality of human life how we make people feel. This is reason in itself to suspect that the way we treat our fellow human beings—the concern and respect we have for one another—are matters having great moral salience.[129] This question is not, of course, one that is addressed by neuroscience; it is instead (like the moral quality of human action) a matter of philosophical debate.[130] Suffice to say for now, if the quality of human lived experience does not have moral significance, it is hard to imagine what would.

Finally, there is the related question of whether people ever have basic

128. *See supra* note 126.

129. In law, notably, the moral significance of consciousness, as a marker of ontological distinctiveness, is implicitly recognized in the modern rule that consciousness (or the capacity to regain it) marks the line between being human and being a corpse, which has a very different legal status. *See, e.g.*, People v. Eulo, 472 N.E.2d 286 (N.Y. 1984); Barber v. Superior Court, 195 Cal. Rptr. 484 (1983).

130. *See supra* text accompanying note 126-2127.

moral equality in the first place—let alone whether it can be affected by their bad deeds. And, alas, as Louis Pojman has argued rather persuasively,[131] none of the current secular arguments for equal human worth is particularly compelling. It is, however, much harder to agree with Pojman's further claim that "there is good reason to believe that humans are *not* of equal worth."[132] Like basic moral equality, moral *in*equality is a strong doxastic claim that requires justification.[133] What is missing from Pojman's claim is any indication of an acceptable basis, logical or empirical, for anyone to assert that he or she is morally superior to any other, privileged to use another as a means or as an object—especially not with all the old-time favorites mostly out of the picture (high birth, race, gender or other "tribal" identity) and with the last holdout, agentic mental causation, now under a dark evidentiary cloud. In short, the situation seems to be one of stalemate, with no compelling argument *for or against* basic moral equality. There simply is no knowable truth of the matter and it is an arrogation to claim that there is. But for the moral egalitarian, this is not such a bad place to be. For if no one can justify a claim of superior moral worth or that others are morally inferior, then the situation is functionally equivalent to basic moral equality—with no one justified to act as though she is elevated or privileged above any other. And we get to this place without the arbitrariness[134] that a blunt presumption of equality would entail. Basic moral equality and the non-existence of demonstrable inequality are, for all practical purposes, the same.[135]

IX. CONCLUSION

Is basic moral equality a permanent human entitlement or a fleeting ephemeron, constantly affected by persons' deeds, good and bad? In particular, do bad actors have less moral worth or deservedness than others,

131. Louis P. Pojman, *Are Human Rights Based on Equal Human Worth*, 52 PHIL. AND PHENOMENOLOGICAL RSCH. 605 (1992).

132. *Id.* at 621 (emphasis added).

133. *Cf.* DAVID O. BRINK, MORAL REALISM AND THE FOUNDATIONS OF ETHICS 100-43 (1989).

134. *See* Pojman, *supra* note 131, at 607-08.

135. Concededly, I am simply assuming that moral normativity, along with moral worth and deserts, do exist in some form (as opposed to not existing at all). This, however, may be said: at the end of the day, the kind of proof we have for the objective reality of morality is essentially the same as what we have for objective physical reality. Our knowledge consists in both cases of inferences drawn from the data that the brain evidently detects in processing energetic impacts on the sensory modalities plus whatever knowledge that may be "wired" into the brain. To be sure, one may always argue that inferences about moral reality are fallacious or false, and such arguments should be considered according to their persuasiveness. But it should never be enough merely to point out that inferences about moral reality have no ultimate foundation, for there are no inferences that do. Both our moral and physical epistemologies each currently depend on a coherent set of inferences within the "local" range of investigation, and that is—for now, at least—the best we expect. For myself, I lean toward a coherentist moral realism of the sort advanced by David Brink. See BRINK, *supra* note 133.

thus justifying the harsh treatment they receive from the criminal justice system? In order to answer these questions in the affirmative, a nexus must be found between persons and their deeds so that the deeds will be suitable criteria for judging the person's moral worth or deserts.

What is commonly thought to make the moral difference and supply the nexus are the mental states—such as intentions, volition, desires or reasons—that, supposedly, cause the body to move. Unlike physical causes of behavior, mental states like intentions and desires do not seem mechanically driven according to physical laws and, therefore, they can be plausibly viewed as "uncaused" by outside forces, *i.e.*, they can be plausibly viewed as the person's own. The problem is there is no evidence for any such "mental causation" except a fallacious and uncorroborated inference from the correlation that is observed between mental states and acts. Mental causation is essentially a speculative basis for inflicting human hardship in the name of criminal justice.

Neural determinism (or neurodeterminism) is the hypothesis that coordinated muscle contractions and, hence, behavior, are caused *only* by neuronal impulses, and neuronal impulses are, in turn, caused *only* by other neuronal impulses and other physical factors, all in accordance with physical laws. As an empirical matter, based on modern neuroscience research, neural determinism is almost certainly true. If so, it would mean that all human behavior is attributable, not to the mental states of the one who acts, but to physical causal chains originating outside the person and acting on her—primarily through the senses. The person-in-action is, in effect, a *conduit* for these causal chains in every move she makes. Since a person's deeds thus cannot be attributed to the mind of the person who does them, they are not suitable as criteria for judging the person's moral worth or deserts.