Washington University
Jurisprudence Review

VOLUME 12 NUMBER 1 2019

DO CRIMINAL MINDS CAUSE CRIME? NEUROSCIENCE AND THE PHYSICALISM DILEMMA

JOHN A. HUMBACH*

ABSTRACT

The idea that mental states cause actions is a basic premise of criminal law. Blame and responsibility presuppose that criminal acts are products of the defendant's mind. Yet, the assumption that mental causation exists is at odds with physicalism, the widely shared worldview that "everything is physical." Outside of law, there is probably no field of secular study in which one can seriously assert that unseen nonmaterial forces can cause physical events. But if physicalism is true then a fundamental premise of modern criminal justice must be false, namely, that criminals deserve punishment because their crimes are the products of their criminal minds.

Efforts to reconcile mind-based theories of criminal responsibility with physicalism encounter a dilemma: how can one say that everything occurs in accordance with physical laws while insisting that offenders deserve blame and punishment precisely because their conduct is not dictated by physical laws? The dilemma highlights the fact that, even if mental states can cause actions, they would have to also be "autonomous" of the physical (and physical laws) to be morally significant. Unless causative mental states are untethered to underlying neuronal activity, people's actions are not their "own" but are merely products of causal chains originating outside themselves far back in time. Thus, if mental states are not autonomous, they could no more have independent moral significance than mental states that are purely epiphenomenal. But the supposition that mental states are autonomous leads to the unpalatable suspicion that the law sends people to prison and deprivation based on spooky-spectral New Age nonsense that no modern thinker would, in any other context, believe.

TABLE OF CONTENTS

INTRODUCTION

A fundamental dilemma is faced by traditional legal thinkers who claim to accept physicalism—a wholly material world without spirits or spooks[1]—but still want to believe that mental states, such as intentions,[2] play a role in deciding what people do. The dilemma concerns the law's requirement that criminal acts be products of the defendant's mind, and not just of the physical body.[3] Our whole system of blame is founded on the belief that mental states can cause actions.[4] People are not ordinarily considered guilty of crimes if their mental states were in no way involved in producing the bodily movements that constituted the criminal act.[5] A reflex, for example, is not a crime, nor are acts done while unconscious.[6] This principle is explicit in the so-called "voluntary act" requirement, and is implicit the various requirements of mens rea that apply to most serious crimes.[7] If, however, we live in a wholly material physicalist world, so that volition and intentions are simply the products of physical causes and effects, it would mean that a fundamental premise of modern criminal justice must be false—namely, the belief that criminals deserve punishment because their acts are the products of their criminal minds.

---

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

1. At its core, physicalism is the view that "everything is physical" (apart from abstract concepts, such as numbers) and that all the concrete stuff of the Universe consists of physical matter and energy, and of nothing else. *See generally* Daniel Stoljar, *Physicalism*, *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed., 2015), https://plato.stanford.edu/entries/physicalism/#TokTypPhy [https://perma.cc/QF8J-M7R6]; *See also* JAEGWON KIM, PHILOSOPHY OF THE MIND 211 (1998) ("[T]he view that there are no concrete substances, in the space-time world other than material particles and their aggregates.").

2. The term "mental state" is used here, as in law, with its "ordinary language, common sense" meaning. *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 (2008) [hereinafter Morse, *Folk Psychology*]; *see also* Stephen J. Morse, *The Inevitable Mind in the Age of Neuroscience* 35, in PHILOSOPHICAL FOUNDATIONS OF LAW AND NEUROSCIENCE (Dennis Patterson & Michael S. Pardo eds., 2016) [hereinafter Morse, *Inevitable Mind*].

3. *See* discussion and authorities cited in the companion to this article: John A. Humbach, *Neuroscience, Justice and the "Mental Causation" Fallacy*, 11 WASH. U. JUR. REV. 191, 195-97 (2019) (hereinafter *Mental Causation Fallacy*).

4. *Cf.* HUME, *infra* note 27.

5. *See, e.g.*, Morissette v. United States, 342 U.S. 246, 250-51 (1952) ("The contention that an injury can amount to a crime only when inflicted by intention is no provincial or transient notion. It is . . . universal and persistent in mature systems of law . . . . A relation between some mental element and punishment for a harmful act is almost . . . instinctive"). *See generally* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 117–44 (6th ed. 2012); WAYNE R. LAFAVE, CRIMINAL LAW 252–88 (5th ed. 2010).

6. *See* MODEL PENAL CODE §2.01(1); *see generally, Mental Causation Fallacy*, *supra* note 3, at 195-97.

7. *See Mental Causation Fallacy*, *supra* note 3, at 195-97.

The difficulty for the avowed physicalist is that legally relevant mental states such as intentions do not seem to be physical; they certainly do not seem to be comprised of the kind physical matter that occupies locations in time and space, nor are they like any form of physical energy found anywhere else in the universe. Rather, the law seems to understand mental causation to operate like some kind of intrabody psychokinesis, the pseudo-scientific process by which just thinking about things can somehow make them happen.[8] While perhaps not as far-fetched as preternatural spirits and spooks, the law's pre-scientific "folk psychology" of mental causation[9] is hard to reconcile with a non-romanticized, evidence-based physicalist view of the world.[10]

One way to try to avoid this dilemma is to re-conceptualize the criminal law's various required mental states as *physical* elements of the crime. Neuroscience provides some support for this approach. It is now essentially undisputed that the "mind" and mental states are dependent for their existence and content on the physical (neuronal) activity of the brain: if the brain is dead, the mind is dead—at least as far as we know. Moreover, it seems clear that all of the content of mental states (viz. mental images, the sounds and voices in our heads and, ultimately, our thoughts and reasoning about them) is ultimately sourced in the information about the world that is collected as raw data by the physical senses and neuronally processed by the brain. The mind has no known independent access to information about the outside world; all such content depends on the brain. And there is no evidence that mental states, such as intentions, can appear in consciousness in the absence of the underlying patterns of neuronal activity with which they seem invariably correlated.[11] And so it could be argued that, just

8. *See Mental Causation Fallacy*, *supra* note 3, at 224-27.

9. *See* Morse, *Folk Psychology*, *supra* note 2, at 1, 2-3, 10-11 ("Roughly speaking, the law implicitly adopts the folk-psychological model of the person, which explains behavior in terms of desires, beliefs and intentions.").

10. Stated another way, the law appears to follow a distinctly *substance dualist* conception of mental states and mental causation. Substance dualism, as "[a]pplied to the present context, posits the existence of both physical bodies and nonphysical minds whereas physicalism admits the existence only of physical objects and processes." JAEGWON KIM, PHYSICALISM, OR SOMETHING NEAR ENOUGH 156 (2005) [hereinafter KIM, PHYSICALISM]. The historically popular but now generally discredited theory of "substance dualism" advanced by Descartes, viz. that the body and mind (or "soul") are two different substances, is most likely the conception of dualism that comes closest to that which is implicit in the law. *See infra* notes 77-78 and accompanying text for further discussion of dualism.

11. *See, e.g.*, R.S. Calabrò, A. Cacciola, et al, *Neural Correlates of Consciousness: What We Know and What We Have to Learn!*, 36 NEUROLOGICAL SCI. 505 (2015), DOI: 10.1007/s10072-015-2072-x, https://www.ncbi.nlm.nih.gov/pubmed/25588680 [https://perma.cc/B2UL-RY7X]; Stanislas Dehaene, et al. *The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications* (2005),

maybe, intentions and other mental states are in fact nothing more than neural activities under a different description—just patterns of physical brain activity that we happen to experience as conscious awareness.[12] If this view of mental states were correct—if mental states are fundamentally *identical* to physical states so it is, indeed, all just physical—the dilemma could be avoided. Physicalist-minded legal theorists could avoid the suspicion that the law sends people to prison and deprivation based on spooky-spectral New Age nonsense[13] that no modern thinker would, in any other context, believe.

Alas, however, the idea that mental states are identical with or otherwise "reducible" to physical states encounters a host of other difficulties[14] and, for reasons suggested below[15] (among others), there is good reason to doubt that our vivid subjective experience of mental states is "nothing more than" just arrangements or interactions of physical matter. The discussion that follows will focus on a difficulty that is of particular legal concern, namely: if the events we call mental states are physical in nature (that is, ontologically physical), then their occurrence, progression and content—everything that we think, intend, desire, reason, etc.—would all be determined in accordance with physical law, and that would deprive mental causation of its distinctive moral significance.

It is central to the physicalism thesis that every event must have either no cause at all (as in the case of random quantum events) or a physical cause that is sufficient, in itself, to produce the event. That is to say, physical events do not "just happen" (except at the quantum level) unless there are sufficient physical causes to make them happen. This principle is called the principle of "Causal Closure"[16] and it is a corollary of physicalism itself: "everything is physical" entails that there are no paranormal or preternatural forces to make events occur.

When applied to mental states (re-conceptualized as essentially physical), Causal Closure would mean that, whenever there are sufficient causes to produce a particular mental state, the mental state has to occur

---

https://www.cs.helsinki.fi/u/ahyvarin/teaching/niseminar4/Dehaene_GlobalNeuronalWorkspace.pdf [https://perma.cc/HLZ9-2RSK].

12. *See* JOHN R. SEARLE, MIND: A BRIEF INTRODUCTION 210 (2004), further discussed *infra* text accompanying notes 58-60.

13. For example, intrabody psychokinesis. *See supra* note 8.

14. An excellent and relatively accessible analysis of these difficulties is provided in KIM, PHYSICALISM, *supra* note 10, at 93-148.

15. *See infra* text accompanying notes 93-96. *See also* text accompanying notes 82-83.

16. More formally, the causal closure of the physical domain holds that "[i]f a physical event has a cause at [time] *t*, then it has a physical cause at *t*." KIM, PHYSICALISM, *supra* note 10, at 15.

automatically; if sufficient causes are absent, the mental state cannot occur. But this means in turn that even if, by physicalizing mental states, we can extricate mental causation from the suspicion of being a pre-scientific fabrication, such a re-conceptualization would deprive mental causation of the independent moral significance it needs to serve as a basis for attributing blame and criminal responsibility.[17] For if mental states are essentially just physical in nature, totally dependent for their existence and content on sensory inputs processed according ordinary physical laws, then both mental states and the actions they produce would no longer be logically attributable to the person who acts. They would be, instead, the inevitable products of chains of causation originating outside the person far back in time. Thus, even if mental causation were true in the sense that physicalized mental states could cause actions, that fact would not logically provide a basis for saying that the criminal defendant's actions are his own.

In sum, the physicalism dilemma is this: The contents of causative mental states (e.g., intentions) can have distinctive moral significance only if they are independent and autonomous from the brain's mechanistic physical processes. But if, as physicalism insists, every physical event that occurs must have a sufficient physical cause, then physicalized mental states would not be independent and autonomous but, rather, would be determined in accordance with physical laws. Such mental states would be, in essence, mere conduits for physical forces borne in chains of causation originating outside the person and, as such, lack independent moral significance. Thus, one must make a choice between either physicalism or mental-state moral autonomy and significance. To claim to embrace both is to land in a contradiction.

None of this made much difference before the advent, in the past few decades, of intensive neuroscience research on the workings of the brain and the physiological bases of behavior.[18] As long as mental causation was the only available plausible explanation to connect the thoughts people have with the things that they do, its reality could go unexamined and unquestioned. Now, however, neuroscience has presented us with a robust

17. Remember, to be criminal, an act must be a product of the defendant's mind, not just the physical body. *See Mental Causation Fallacy, supra* note 3.

18. A recent study found that 1.79 million articles in the area of brain and neuroscience research were published from 2009 to 2013. Georgin Lau et al., *New Report Maps the Landscape of Global Brain Research*, ELSEVIER CONNECT (Nov. 13, 2014), https://www.elsevier.com/connect/new-reportmaps-the-landscape-of-global-brain-research [https://perma.cc/5G8X-PTXB]; *see also* RICHARD PASSINGHAM, COGNITIVE NEUROSCIENCE: A VERY SHORT INTRODUCTION 3 (2016) ("[N]early 30,000 experiments conducted using fMRI alone."). Much of this research is described and summarized in ROBERT M. SAPOLSKY, BEHAVE: THE BIOLOGY OF HUMANS AT OUR BEST AND WORST (2017).

and detailed physiological explanation of how the behavior of people (and all other organisms with brains) is generated. According to this explanation,[19] the brain neuronally processes information that comes in through the senses together with physically embodied information from previous inputs (memories) to determine the coordinated outputs that are sent through motor neurons to the muscle fibers that produce the bodily movements that constitute human behavior. The neuroscience explanation of how behavior comes about leaves no place for mental causation to play a role; it contains no gaps that mental causation is needed to fill.[20] By contrast, mental causation—if it exists at all—remains completely unexplained, a naked conjecture based almost solely on logically fallacious inferences from blunt correlations that have well-documented physical explanations.[21] In light of the findings of neuroscience it is extremely improbable that there is such a thing as mental causation.[22]

There has, however, emerged a resilient corps of neuroskeptics who resist the conclusion that mental states play no part in producing human behavior and deny that neuroscience requires a broad rethinking of criminal justice practices. As pointed out by Professor Stephen J. Morse, a leading proponent of the neuroskeptical view, neuroscience's mechanistic explanation of behavior is at odds with the longstanding and deeply-entrenched belief that people have "agency," viz. the ability to act intentionally and for reasons as the authors of their own acts.[23] And moral responsibility, he adds, depends on agency.[24] But Professor Morse and his

19. *See* SAPOLSKY, *supra* note 18, at 21–77, 535–36 (describing and summarizing how the brain chooses and produces bodily movements); PASSINGHAM, *supra* note 18, at 66–81; and other sources cited in *Mental Causation Fallacy, supra* note 3, at 193 n.6.

20. This is not to say there is not still much to learn about the details. Moreover, a frustrating reality is that, because the production of human behavior is such a radically multifactorial process, the task of getting and reading of all the many factors and taking them into due account is beyond present technologies. But it by no means follows that, just because we cannot measure all the physical factors that affect behavior, we are therefore free to assume that there are non-physical factors at work. *See Mental Causation Fallacy*, *supra* note 3 at 199-200. The weather and the stock market are similarly multifactorial, but nobody assumes for that reason that they are influenced by forces from outside the physical domain.

21. *See* SAPOLSKY, *supra* note 18, at 21-77, 535-36 (describing and summarizing how the brain chooses and produces bodily movements); PASSINGHAM, *supra* note 18, at 66-81; and other sources cited in *Mental Causation Fallacy, supra* note 3, at 193 n.6 and 224-38.

22. *See Mental Causation Fallacy, supra* note 3, at 216-34, 238-43.

23. Stephen J. Morse, *Criminal Law and Common Sense: An Essay on the Perils and Promise of Neuroscience*, 99 MARQ. L. REV. 40, 67 (2015) [hereinafter Morse, *Common Sense*] (describing agents as "creatures who act intentionally for reasons, who can be guided by reasons, and who in adulthood are capable of sufficient rationality to ground full responsibility unless an excusing condition obtains"). *See also infra* text accompanying notes 66-67.

24. *Id*. at 67 ("no responsibility is possible if we are not agents"); Morse, *Inevitable Mind*, *supra* note 2, at 36 ("responsibility judgments depend on … mental states"). *See also infra* text

fellow neuroskeptics face the dilemma just described: The mental causation hypothesis cannot be reconciled with physicalism unless mental states are indeed dependent on and controlled by the neuronal activity that underlies them. But if they are thusly dependent and controlled, they would lack the independent moral significance that is needed for ascribing responsibility and blame.

The discussion that follows will expand on these points. Part I discusses the fundamental contradiction that is encountered when trying to embrace both physicalism as well as the idea that human beings are "agents" who are not subject to the usual constraints described in physical laws. Part II considers whether either "property dualism" or "emergentism" can provide an avenue to escape that contradiction, concluding they cannot. Part III notes that, while it does not seem plausible that mental states are fully "reducible" to physical states (that they are "nothing more" than physical states), that fact has no bearing on whether mental states have causative properties of their own. Finally, Part IV briefly discusses why it is important to criminal justice policy that we recognize the fundamental contradiction that infests the neuroskeptical view.

I. MINIMAL PHYSICALISM AND "AGENCY": THE CONTRADICTION

When deciding cases, assessing damages, sending people to prison and otherwise resolving disputes, courts are normally meticulous in ascertaining the facts on which their judgments are based. However, there is a notable exception in the case of mental causation.[25] The requirement that, for most serious crimes, the defendant is criminally responsible only if she intentionally, knowingly or recklessly engaged in the forbidden conduct or caused the forbidden result.[26] In assessing criminal guilt the law simply assumes that, if the defendant acted with a culpable mental state, the wrongful act was the product of the defendant's culpable mind.[27] However,

---

accompanying notes 40-48. *See generally* Markus Schlosser, *Agency*, *The Stanford Encyclopedia of Philosophy,* (Edward N. Zalta ed., 2015) https://plato.stanford.edu/entries/agency/#SenAge [https://perma.cc/42AE-XAA3] ("The philosophy of action … construes action in terms of intentionality, while [the standard theory of action] explains the intentionality of action in terms of causation by the agent's mental states and events. From this, we obtain a standard conception and a standard theory of agency.").

25. *Mental Causation Fallacy*, *supra* note 3.

26. *See supra* note 5 and accompanying text.

27. The *presence* of culpable mental states must be proved, as well as their concurrence with the crime's actus reus, *see* State v. Rose, 311 A.2d 281 (R.I. 1973), but courts seem never to discuss whether there is proof that the required mental states actually *caused* the criminal acts to occur. While, as David Hume famously observed, causation itself is never directly observable, DAVID HUME, AN ENQUIRY CONCERNING HUMAN UNDERSTANDING 60 [Sec. VII Pt. I ¶ 6] (1988) (1748), Hume's

the putative causal efficacy of mental states does not even purport to be based on known facts about physical reality.

There is probably no field of secular study in which a scholar can seriously assert that unseen immaterial forces are the true cause of physical events and not be laughed away. It is therefore unsurprising that the neuroskeptic friends of mental causation are eager to claim the mantle of physicalism,[28] assuring their readers that they, too, believe "we inhabit a thoroughly material, physical universe in which all phenomena are caused by physical laws."[29] This avowed commitment to physicalism appears, however, to present neuroskepticism with a dilemma. How can one accept that "all phenomena are caused by physical laws"[30] and then, at the same time, also insist that offenders deserve blame and punishment precisely because their wrongful conduct is caused by mental states, like intentions and reasons,[31] that are *not* shackled to or determined by the mechanical laws of physics? [32]

Physicalism, or materialism, is a view of the world that rejects the "magical thinking" of the past in favor of explanations of events that are based on non-fallacious inferences from reliable physical evidence. Strictly speaking, it takes no position on matters as to which physical evidence is absent, but it insists that, where events have robust evidence of ordinary physical causation, non-physical explanations that lack supporting evidence

---

observation does not mean that causation cannot be *inferred* from circumstantial evidence. It can; we do it every day, and some inferences are stronger than others, depending on the evidence. *See Mental Causation Fallacy*, *supra* note 3 at 238-43 ("inference to best explanation"). But the inference of mental causation is particularly weak and, indeed, fallacious, *id.* at 224-34, far too weak to serve the foundation and rationale for ascriptions of criminal responsibility, blame and guilt. Nor will it do to say (as some have tried) that the mere *presence* of culpable mental states is enough to justify ascribing blame and criminal responsibility, whether or not a defendant's mental states have actual causal efficacy. *See, e.g.*, *Id.* at 196 n.16. If a defendant's culpable mental states are causally inert and, thus, have nothing to do with whether the crime occurs or not, their presence or not at the time of the crime should be legally irrelevant and it would be prejudicial error to inform the jury of them. It is like punishing a person for an earthquake because he ardently wished that it would occur.

28. The basic premise of physicalism is that all of the concrete stuff of universe consists of physical matter and energy, and of nothing else. *See supra* note 1.

29. *See* Morse*, Common Sense*, *supra* note 23, at 47; Morse, *Inevitable Mind*, *supra* note 2, at 33.

30. *Id.*

31. *See, e.g.,* Morse, *Inevitable Mind*, *supra* 2, at 33-34 ("Virtually everything for which we deserve to be praised or blamed and rewarded or punished is the product of mental causation. . . .The brain, in contrast, is a machine, an intricate, organic electrochemical machine. . . [M]achines may cause harm, but they cannot do wrong . . . . [O]nly people can do wrong. Machines do not deserve praise, blame, reward, or punishment.").

32. *See* quotations in preceding footnote.

must be eschewed. Physicalism is, in other words, the convention[33] of strongly preferring physical explanations as the lingua franca or common ground for explaining and relating the events of the world. Its practical justification, based on centuries of experience, is that unless we require reliable physical evidence as the test for what to believe, there would be no end to the proliferation of imagined possible hypotheses and no dispositive way to decide among them. While all should be entitled to the beliefs they find personally comforting ("non è vero ma ci credo," as the Neopolitans say[34]), it is another matter entirely for the state to punish and deprive on the basis of "truths" that are merely subjective.

On the surface, the whole idea of mental causation seems at odds with even a "minimal" physicalism,[35] which entails that "what happens in our mental life is wholly dependent on, and determined by, what happens with our bodily processes."[36] To be sure, minimal physicalism does not rule out the possibility of a purely parallel mental causation, a kind of "mental" causation that merely piggybacks on neuronal causation and whose effects on behavior are indistinguishable from the brain's.[37] But such a duplicative and functionally superfluous mental causation—which merely rides along with physical causation—hardly seems to count as the kind of "causation by intentions and reasons" that could provide the "agency," which is said to be requisite for blame and punishment.[38] All actions putatively caused by

33. That is, it is not an article of faith nor is it an ontological certainty, though it is probably as near to one as any empirical matter can be. *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/5A59-5GG9] (presenting arguments that have led to the near-universal acceptance of physicalism among modern scholars and researchers).

34. "It's not true, but I believe it."

35. Jaegwon Kim uses the term "minimal physicalism" to refer to "the shared minimum commitment of all positions that are properly called physicalist." KIM, PHYSICALISM, *supra* note 10, at 13.

36. KIM, PHYSICALISM, *supra* note 10, at 14.

37. *See generally* KIM, PHYSICALISM, *supra* note 10, at 32-46. Although minimal physicalism seems to include a principle of "Causal Closure" (roughly, every event has physical causes if it has any cause at all), *see* KIM, PHYSICALISM, *supra* note 10, at 15-16; David Papineau, *The Causal Closure of the Physical and Naturalism* 53, *in* THE OXFORD HANDBOOK OF PHILOSOPHY OF MIND (Beckermann, et al. ed., 2009); and *supra* notes 16 and accompanying text, the principle of Causal Closure does not in itself necessarily rule out mental causation. That is, minimal physicalism and Causal Closure do not in themselves deny the possibility that non-physical causes can co-occur with physical ones, as long as the physical causes would be sufficient causes in themselves. *See* KIM, PHYSICALISM, *supra* note 10, at 17. The problem for piggybacking mental causes is created, not by Causal Closure but by the so-called "exclusion principle." *See infra* note 38.

38. *See* Morse, *Inevitable Mind*, *supra* note 2, at 46; Morse, *Common Sense*, *supra* note 23, at 67; *see also* Pardo & Patterson, *infra* note 49, at 16 [online version] (making the same point). What is more, such a "piggybacking" variety of mental causation would seem to be excluded by the so-called "exclusion principle," which states that when there are two apparent causes of an event and one is wholly

that sort of parallel mental causation would, like the neuronal causation on which it piggy-backs, be mechanistically determined in accordance with physical laws.

The would-be physicalist legal thinker seems therefore to be faced with an unpalatable choice: either renounce physicalism and its insistence that human behavior is determined by mechanistic physical laws, or embrace physicalism and give up the idea that mental states can play the distinct, non-mechanical role in causing conduct that is supposedly crucial to moral responsibility and punishment.[39]

The dilemma can be illustrated by considering a contradiction that lies at the foundation of the neuroskeptical thesis, which will be represented here as laid out in the writings of Professor Morse, a leading and thoughtful exponent of the neuroskeptical viewpoint.[40] When specifying his premises, Professor Morse insists that he accepts physicalism[41] and he agrees that "human action and consciousness are *produced by the brain*, a material organ that works according to biophysical laws."[42] In other passages, however, he seems less convinced. He writes, for example, that mental states are "fundamental to a full explanation of human action,"[43] that we do what we do "*because we intend* for reasons,"[44] that mental states "*produce* … bodily movements,"[45] and that there cannot be "genuine responsibility" unless people "have the capacity to *determine* their actions by reasons and to act in light of those reasons."[46] He repeatedly refers to the major moral

---

dependent on the other, the dependent cause is otiose and "excluded" by the cause on which it depends. *See generally* KIM, PHYSICALISM, *supra* note 10, at 32-69. Kim's statement of the exclusion principle is as follows: "No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination [viz. causation of an event by two distinct causal chains]." *Id.* at 42. The point of exclusion principle is to avoid an arbitrary proliferation of putative causes. For more on the exclusion principle and Professor Morse's attempt to dismiss it, *see infra* note 62.

39. Morse, *Inevitable Mind*, *supra* note 2, at 33 and *infra* notes 58-60; *see also* other quotations *infra* text accompanying notes 43-49, 69-70.

40. *See, e.g.*, Morse, *Common Sense*, *supra* note 23; Morse, *Inevitable Mind*, *supra* note 2.

41. *See, e.g.*, Morse, *Common Sense*, *supra* note 23, at 47 ("I am a physicalist"); Morse, *Inevitable Mind*, *supra* note 2, at 33. Specifically, in the mind-body context, he says physicalism means that "[t]he brain enables the mind and action, but we have no idea how." Morse, *Common Sense*, *supra* note 23, at 47.

42. Morse, *Inevitable Mind*, *supra* note 2, at 33 (emphasis added).

43. Stephen J. Morse, *Lost in Translation: An Essay on Law and Neuroscience* 530 (2011), *in* LAW & NEUROSCIENCE (Michael Freedman ed., 2011) [hereinafter Morse, *Translation*].

44. Morse, *Inevitable Mind*, *supra* note 2, at 33 (emphasis added).

45. *Id.* ("Responsibility judgments depend on the mental states that produce and accompany our bodily movements") (emphasis added).

46. Morse, *Common Sense*, *supra* note 23, at 48 (emphasis added). Although Professor Morse states that "contra-causal freedom is simply not necessary" for responsibility, Morse, *Inevitable Mind*, *supra* note 2, at 45, this statement appears only to reaffirm his provisional acceptance of determinism in

difference he sees between "persons" and "machines"—for example, when he insists that an "electrochemical machine" like the brain cannot be a responsible agent or do wrong because "only people can do wrong,"[47] or when he notes approvingly that "[t]he law treats persons generally as intentional creatures and not simply as mechanistic forces of nature."[48] Assertions like these can only mean that there is, in Professor Morse's view, a lot more to "agency" and human behavior than just doing whatever is dictated by the physiological brain. On the contrary, despite his avowed commitment to physicalism, he unambiguously ascribes a distinct causal role to mental states and sees them as able to rise above and transcend the "electrochemical machine" in producing human behavior.[49] He does not, in short, sound at all like someone who actually believes in a thoroughly material, physical Universe.

The direct contradiction in Professor Morse's neuroskeptical thesis is hard to ignore. It cannot be true (as physicalism insists) that "all phenomena are caused by physical laws"[50] and also be true that we are "creatures whose desires, beliefs, and intentions play a causal role in explaining our behavior"[51]—at least they cannot be playing a *distinctive* (non-piggybacking) causal role. It is a contradiction that seems, moreover, to be intrinsic to any *compatibilist*[52] theory of responsibility that builds on the

---

general, and it does not say (nor is it apparently meant to say) that responsibility can exist even if human behavior is entirely dictated by the biomechanical activity of neurons.

47. Morse, *Inevitable Mind*, *supra* note 2, at 33-34; Morse, *Folk Psychology*, *supra* note 2, at 6.

48. Stephen J. Morse, *Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty,* 62 MERCER L. REV. 837, 840 (2011).

49. *Cf.* Iskra Fileva & Jonathon Tresan, *Will Retributivism Die and Will Neuroscience Kill It?* COGNITIVE SYSTEMS RESEARCH 9 (2015) (describing, but not necessarily endorsing, a similar conception of "agent-causation" which supposedly reconciles "full causation" with agent autonomy, reasoning that full causation occurs, as long as something caused the existence of the agent even if the agent, thus caused to exist, possesses the freedom to "make things happen" that are "caused by her" in the sense of *determined* by her).

50. Morse, *Common Sense*, *supra* note 23, at 47; Morse, *Inevitable Mind*, *supra* note 2, at 33.

51. Stephen J. Morse, *Scientific Challenges to Criminal Responsibility*, in JOEL FEINBERG ET AL., PHILOSOPHY OF LAW 844 (9th ed. 2014).

52. Compatibilism is (as described by Professor Morse) "a set of similar theories that hold with varying intensity that responsibility is possible in a deterministic universe as long as agents have the capacity to act according to their reasons" and intentions. Morse, *Inevitable Mind*, *supra* note 2, at 45; Stephen J. Morse, *NeuroEthics: NeuroLaw* OXFORD HANDBOOK ONLINE (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2919011 [https://perma.cc/T4DV-ZEZP]. *See generally* Michael McKenna & D. Justin Coates, *Compatibilism, The Stanford Encyclopedia of Philosophy,* (Edward N. Zalta ed., Winter 2018), https://plato.stanford.edu/archives/win2018/entries/compatibilism/.

assumption that human behavior is causally controlled or influenced by reasons, intentions, beliefs, or other mental states.[53]

What makes the contradiction particularly intractable is the fact that mental-state causation would have no distinctive moral significance unless the causative mental states at work are, at least in some degree, *autonomous* of the neuronal activity that co-occurs with them. That is to say, in order for mental causation to make a moral difference, the processes of mentation that generate intentions, reasons and other causative mental states would have to be free to wander along paths of their own, through the byways of reason and thickets of thought, their content not tied to or dictated by the computational processes of the biomechanical brain.[54] If causative mental states and the streams of thought leading up to them are not autonomous—if they simply mirror the activity and outputs of the computational brain—then mental causation would be essentially just a conduit for brain-state causation and wrongdoers would be blameless "victims of neuronal circumstances."[55] To avoid this conclusion and hold that mental states make

---

53. Michael Pardo and Dennis Patterson put this contradiction into sharp focus in summarizing Professor Morse's position: "In sum, mental states are produced by and realized in the brain (they are not a distinct substance), but they nonetheless exist and play a (non-reductive) causal role . . . ." Michael S. Pardo & Dennis Patterson, *Morse, Mind and Mental Causation*, 11 CRIM. L. & PHIL. 111, 113 (2017) (describing their critique as "largely a friendly one"). In other words, what Professor Morse seems to be saying per Pardo and Patterson is in effect that mental states are physical (not a non-physical substance) and mental states are irreducibly non-physical—a logical contradiction. The expression "P and not-P" is a contradiction. *See* Tom Ramsey, A BRIEF OVERVIEW OF LOGIC, http://www.math.hawaii.edu/~ramsey/Logic/PandNotP.html [https://perma.cc/JWA3-LC8K]

54. I am tempted to say that causative mental states can make a moral difference only if they are "anomalous" in a Davidsonian sense. *See* Steven Yalowitz, *Anomalous Monism*, *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed., Winter 2014), https://plato.stanford.edu/archives/win2014/entries/anomalous-monism ("There are no strict laws on the basis of which mental events can predict, explain, or be predicted or explained by other events."). It is probably best, however, to avoid the baggage that might come with adopting Davidson's conception.

Another way of saying that mental states are autonomous of brain states is to say that they are not "supervenient" on brain states—that is, the properties of mental states are independent of, and not necessarily co-variant with, the properties of their underlying brain states. *See* Brian McLaughin & Karen Bennett, *Supervenience, The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed., Winer 2018), https://plato.stanford.edu/entries/supervenience. 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). The reason for my taking this view is that supervenience seems to be required by physicalism (or, more specifically, the Causal Closure principle discussed *supra* note 16 and accompanying text and *supra* note 37). However, I have never seen reference to proof that mental states are *necessarily* supervenient as a matter of empirical fact and, accordingly, I think it better not to raise supervenience as a reason for questioning the credibility of the mental causation hypothesis. In other words, I am assuming for purposes of argument in the text that mental states *might* be autonomous in order to follow out the implications of saying they are and to show the dilemma that such implications pose for an avowed physicalist stance.

55. *See Mental Causation Fallacy*, *supra* note 3, at 210-12. *See also* Michael S. Pardo & Dennis Patterson*, Morse, Mind and Mental Causation,* 11 CRIM. L. & PHIL. 111, 121 [16 online version]

a moral difference, a way must be found to theorize mental states as unshackled from their physiological base. Mental states would have to be able to produce other mental states as a person's stream of consciousness flows along, and they would have to be able to produce physical impulses in the motor neurons (that signal the muscles to contract) so that the results of the mind's reasoning could be realized as physical actions and behavior.[56] In short, if it is true that mechanistically-determined neuronal causation of wrongdoing alone does not justify infliction of punishment, the addition of parallel, piggy-backing mental causation could not justify it either. For mental causation to have a moral significance, the causative mental states have to be autonomous and independent of the neurophysiological substrate, not mere "Pinocchios" dominated by the brain.[57]

For example, the totally brain-bound (and therefore non-autonomous) mental causation described by John Searle would *not* provide a suitable basis for justifying attributions of responsibility or blame.[58] According to Professor Searle,

> the mental is simply a feature (at the level of the system) of the physical structure of the brain . . . . [C]ausally speaking, there are not two independent phenomena, the conscious effort and the unconscious neuron firings. There is just the brain system, which has one level of description where neuron firings are occurring and

---

(2017) (making essentially the same point and suggesting that Professor Morse needs an alternative to mental causation as the basis for ascribing responsibility).

As further described in *Mental Causation Fallacy*, *supra* note 3, at 211-12, the "victim of neuronal circumstances" hypothesis comes from a much-noted article by Joshua Greene and Jonathan Cohen. Joshua Greene & Jonathan Cohen, *For the Law, Neuroscience Changes Nothing and Everything*, PHIL. TRANS. ROYAL SOC'Y 1775–1785 (2004). In that article, Greene and Cohen pointed to the considerable evidence that everything a person does is determined by biomechanical neuronal processes, with wrongdoers haplessly acting out the prescribed script of fate.

56. Stated another way, the mind would have to be able to pursue sequences of reasoning from given starting points to conclusions that are not the same as the conclusions that are reached by the biomechanical computations of the brain. The mind would have to be able, moreover, to inject its behavioral choices into the appropriate pre-motor brain areas so the motor neurons would be caused to activate the muscles accordingly.

57. *See* Morse, *Inevitable Mind, supra* note 2, at 49 ("We are not Pinocchios and our brains are not Geppettos pulling the strings."). It is said to be "widely accepted" that mental states are *not* autonomous and that there are "lawful [nomological] correlations between sensory experiences and physical/functional states" of the brain. KIM, PHYSICALISM, *supra* note 10, at 126-27. The point here in the text is not, however, to take a position on that "widely accepted" view or to invoke it as an argument against mental-state autonomy. Rather, the only point being made in the text is that one would have to forsake physicalism in order to embrace such autonomy, since the two are inconsistent.

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

> another level of description, the level of the system, where the system is conscious and indeed consciously trying to raise its arm.[59]

If Professor Searle is correct that behavior-causing mental states and physical brain states are just two "levels of description" of the same thing, then the "two" would have exactly the same causal effects and, accordingly, would produce the exact same behavior. As long as we assume (with Searle) that neuron firings occur in accordance with physical laws, it follows that the successive states of a person's overall "brain system" (and therefore of her successive mental states) are also dictated by physical laws.[60] If that is so, then any mental causation that may appear to exist would not be autonomous but would simply be a mirroring of the physical succession of brain states. As such, it would add no independent factual element, beyond the neuronal, that could justify blaming and punishment. In such a picture it would still be the firing of neurons, not the dependent mental states, that would be "in control" and doing the real causal work. It would, in other words, still be the person's brain states, not her mental states, that are the morally-relevant causes of her wrongs.[61] She would be, in short, a "victim

---

59. *Id.*

60. SEARLE, *supra* note 58, at 114, 209 ("mental processes … have no causal powers in addition to those of the underlying neurobiology" and "there is no causal efficacy to consciousness that is not reducible to the causal efficacy of its neuronal basis").

61. This is essentially, I think, the point that is made by the co-called "exclusion principle," which says, roughly, if an event has a sufficient physical cause and also a mental cause, and the two are not factually independent in their origins, then the physical cause "excludes" the mental cause—or, more precisely:

> If an event *e* has a sufficient cause *c* at time *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination).

KIM, PHYSICALISM, *supra* note 10, at 17. *See generally id.* at 17-22. What this means for the present discussion is, in broad paraphrase:

> When it appears that mental states and the brain states on which they depend are both simultaneously causing certain conduct, and the brain states alone would be sufficient to cause the conduct, then the actual causal work is being done only by the brain states and only they can be properly regarded as the cause.

It should be noted, however, that the exclusion principle does not in itself "disprove" the possibility of *autonomous* mental causation unconditionally, but only does so if there is a sufficient physical cause. In other words, the exclusion principle would only hold that mental causation is impossible if every physical event does indeed either have no cause or have a sufficient physical cause (i.e., if Causal Closure is true, *see supra* note 16 and accompanying text). While it seems likely that Causal Closure is true, one cannot of course disprove autonomous mental causation by *assuming* the truth of Causal Closure without completely begging the question. What the exclusion principle *does* say, however, is that, as long as mental states are not "distinct" from (i.e., not autonomous from) underlying brain states, they cannot be properly considered the causes of conduct, since the brain states would be doing all the causal work.

Professor Morse attempts to brush aside the exclusion principle citing an article that provides, he says, "plausible philosophical reason to believe" that the principle, appropriately reformulated, actually

of neuronal circumstances."[62] If victims of neuronal circumstances are not responsible[63] then Professor Searle, though he has succeeded in maintaining his commitment to both mental causation and physicalism, has done so by conceiving of mental causation in a way that denies it is autonomous and, therefore, denudes it of distinctive moral significance.[64]

The crucial importance of mental-state autonomy to responsibility has not been lost on those who see culpable mental states as a basis of criminal responsibility.[65] Indeed, mental-state autonomy seems to be taken for granted in the folk psychology that underlies the criminal law. As Professor Morse has written, the criminal law "presupposes a 'folk psychological' view [that] causally explains behavior in part by mental states such as desires, beliefs, intentions, willings, and plans,"[66] and it is an "incoherent notion" to have genuine responsibility without the ability to act intentionally, rationally and for reasons.[67] Statements like these recognize that, if the content of mental states is dictated by mechanistic brain states

---

allows for mental causation. Morse, *Inevitable Mind*, *supra* note 2, at 34, *citing* Christian List & Peter Menzies, *Non-Reductive Physicalism and the Limits of the Exclusion Principle,* CVI J. OF PHIL. 425 (2009). It appears, however, that Professor Morse is "overclaiming." Without getting too deep into the weeds, the main problem with relying on the cited article is that it does not (by its own admission) use the definition of "cause" (viz. "generative" or "productive" causation) that is generally accepted in discussions of the exclusion principle. *See* KIM, PHYSICALISM, *supra* note 10, at 18. In consequence, the reasoning of the List & Menzies article slips past the gravamen of the exclusion principle like a ship passing in the night. *See also* Jose Luis Bermudez & Arnon Cahen, *Mental Causation and Counterfactuals*, *in* CONSCIOUSNESS AND THE ONTOLOGY OF PROPERTIES, 155-173 (Mihretu P. Guta ed., 2019) (critiquing the List & Menzies analysis).

62. *See supra* note 55.

63. And Professor Morse contends they are not, due to lack of agency. *See Common Sense*, *supra* note 23, at 67*; Inevitable Mind, supra* note 2, at 46.

64. It appears that the same can be said of Hilary Bok's analysis in *Want to Understand Free Will? Don't Look to Neuroscience*, CHRON. OF HIGHER ED. online (Mar. 18, 2012), https://www.chronicle.com/article/Hilary-Bok-Want-to-Understand/131168 [https://perma.cc/GT9B-4DNE] ("[T]he claim that a person chose her action does not conflict with the claim that some neural processes or states caused it; it simply redescribes it.").

65. Morse, *Inevitable Mind*, *supra* note 2, at 35. *But cf.* Katrina L. Sifferd*, What Does It Mean to Be a Mechanism? Stephen Morse, Non-reductivism, and Mental Causation,* 11 CRIM. L. & PHIL. 143 (2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512325 [https://perma.cc/5TNP-RB3E] (arguing that even if the mental is reduced to the physical "[i]ntentional mental states are real, and they cause action *because they are physical things* in the world that participate in causal relations") (emphasis added). Professor Sifferd does not however explain how, if mental states are "physical things," they could possibly have the distinctly moral implications for responsibility that physical brain states do not.

66. Morse, *Common Sense*, *supra* note 23*,* at 49. *Accord,* Morse, *Inevitable Mind*, *supra* note 2, at 35 ("[R]esponsibility judgments depend on … mental states"); Morse, *Common Sense*, *supra* note 23*,* at 40, 69 ("[N]o responsibility is possible if we are not agents"); and Morse, *Translation*, *supra* note 43, at 531 (the law views "the person as a conscious … creature who forms and acts on intentions that are the themselves the product of the person's other mental states such as desires, beliefs, willings, and plans.").

67. Morse, *Common Sense*, *supra* note 23, at 40, 69.

and the two of them always produce the exact same behavior, then wrongdoers would be in essence "victims of neuronal circumstances" and, as such, not responsible for their misdeeds.[68] To say that brain states are mechanistic is to say that the brain is essentially like a machine and, though "machines may cause harm, … they cannot do wrong [or] deserve … punishment."[69] Therefore, criminal responsibility requires there to be something more to the gubernation of human behavior than just the biomechanical operations of "an intricate, organic electrochemical machine."[70] And the "something more" that is required to make persons criminally responsible cannot just be the "piggybacking" variety of mental causation that does no more than replicate the causal results that the physical brain would produce anyway. There has to be *autonomous* mental causation that is causally independent of the physical and, indeed, able to override the forces of physical causation generated by the brain.[71] But autonomous mental causation flatly contradicts the physicalist premise that physical events cannot be produced by non-physical (immaterial or spectral-spirit) causes.[72] The position that people are responsible and deserve punishment based on mental causation contradicts a commitment to even a minimal physicalism. This contradiction creates serious doubts as to the justness of a criminal justice system that systematically inflicts hardship, deprivation and misery on millions of people on the theory that non-physical mental states cause physical criminal acts.

It is, of course, easy enough to see why theorists like Professor Morse have been driven to this contradiction: Not wanting to place themselves at odds with modern mainstream scholarship, neuroskeptics must eschew the out-of-favor "notion that we have an immaterial mind or soul that is somehow in causal relation with our physical body."[73] But causally efficacious mental states that are operationally autonomous of the physical domain are, by definition, "immaterial." It is not consistent to say both that "we inhabit a thoroughly material, physical universe" and also that we are "creatures who act intentionally for reasons, who can be guided by

68. *See supra* note 55 and accompanying text.

69. Morse, *Inevitable Mind*, *supra* note 2, at 34.

70. *Id*. at 33-34.

71. Morse, *Common Sense*, *supra* note 23, at 49 (stating that "genuine responsibility" can only exist if people "have the capacity to determine their actions by reasons and to act in light of those reasons").

72. Certainly, Professor Morse concedes this truism and perhaps even slightly overstates it. *See* Morse, *Common Sense*, *supra* note 23, at 47 ("[A]ll phenomena are caused by physical laws"); Morse, *Inevitable Mind*, *supra* note 2, at 33. In other words, autonomous mental causation contradicts the principle of Causal Closure. *See supra* notes 16, 36.

73. Morse, *Inevitable Mind*, *supra* note 2, at 33.

reasons."[74] Ardently seeking to avoid the trap of substance dualism,[75] Professor Morse appears in the end to have traded it in for a fundamental contradiction instead.

Understandable or not, however, this contradiction is a considerable problem for any theory of responsibility that tries to embrace both physicalism and autonomous mental causation. Its presence in an argument for responsibility and punishment based on mental causation means the conclusion—that punishment is "deserved"—is not supported by lines of argument that coherently follow from a set of consistent premises. A conclusion that proceeds from contradictory premises is not reliable.[76] By accepting both physicalism as well as immaterial, preternatural causes, the compatibilist/neuroskeptic case for responsibility and punishment can only be deemed inconclusive, at best. Instead of establishing that "mental causation" provides a morally just basis for subjecting millions of people to the hardship and deprivation called punishment, internally inconsistent arguments that stray from the path of physicalism into the murky realm of incorporeal forces only reinforce the conclusion that no such case for punishment can be made.

II. PROPERTY DUALISM AND EMERGENTISM

While the long-dominant (and decidedly non-physicalist) idea that the brain and mind are separate *substances* fell into disfavor during the last century,[77] the view has persisted that the mind and brain are fundamentally different in nature. In place of the old discredited "substance dualism,"[78]

74. Morse, *Common Sense*, *supra* note 23, at 40.

75. Professor Morse, for example, wants it to be very clear that he is not contemplating any form of substance dualism. *See* Morse, *Translation*, *supra* note 43, at 536 ("[I]t may seem, therefore, as if law's emphasis on the importance of mental states as causing behavior is based on a pre-scientific, outmoded form of dualism, but this is not the case.").

76. When a premise is a logical contradiction, everything follows as a logical conclusion from it. *See* Tom Ramsey, A BRIEF OVERVIEW OF LOGIC, http://www.math.hawaii.edu/~ramsey/Logic/PandNotP.html [https://perma.cc/F24Q-4RJJ].

77. KIM, PHYSICALISM, *supra* note 10, at 151.

78. *Id.* at 156 ("Applied to the present context, substance dualism posits the existence of both physical bodies and nonphysical minds whereas physicalism admits the existence only of physical objects and processes.").

new hypotheses known as "property dualism"[79] and "emergentism,"[80] have been proposed to reconcile mental causation with physicalism. The basic strategy of these hypotheses is to agree that there is only one kind of "substance," namely physical (hence, physicalism is true), but to argue nonetheless that mental states can exist as something distinct from the physical because salient non-physical properties can "emerge" out of arrangements and systems of physical stuff.[81] For example, a recognizable image of Mickey Mouse can "emerge" out of a bunch of glowing pixel dots on a computer screen. Even though the image is physically just a collection of luminescent dots, it can have aesthetic and symbolic properties that are "over and above" the physical properties of the dots themselves. Because the emergent image has non-physical properties that are over and above the physical properties of its constituents, the situation can be referred to as a case of "property dualism"; there is only one substance (physical) but it has two kinds of properties, physical and non-physical.

In the mental-causation context, the point of property dualism and emergentism is to allow one to disavow the "substance dualism" idea that the mind and body are different substances[82] while still maintaining that mental states are different and distinct from their underlying physical base and, as such, are potentially able to have causal properties that are independent of the physical. For example, one could claim that, even though a particular pattern of synaptic discharges is physical in nature, the pattern of discharges can have non-physical properties by being experienced in consciousness as, say, the redness of cherries, the pain of a pinprick or an

79. Property dualism is the idea that physical events can have non-physical properties that cannot be fully explained by the properties of the underling physical substrate itself. *See* KIM, PHYSICALISM, *supra* note 10, at 20–22, 156–61; Howard Robinson, *Dualism*, *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed., Fall 2017), https://plato.stanford.edu/archives/fall2017/entries/dualism/. Unlike substance dualism, property dualism accepts that there is only one substance—the physical—but, as explained in the text, it finds hope for mental causation in the idea that physical substances, such as arrangements of atoms, can have non-physical properties, such as consciousness.

80. Emergentism is a form of property dualism which holds that certain arrangements of atoms can give rise to discernible objects and properties whose behavior cannot be entirely accounted for by the properties of the atoms themselves. For example, arrangements of dots can be viewed as a picture that has properties over and above the properties of the dots themselves, or certain arrangements of atoms or neural firings can give rise to conscious states with properties over and above those of the atoms or neural firings themselves. KIM, PHYSICALISM, *supra* note 10, at 157–58 ("In addition to physical properties [of the substance of the brain], there are physically irreducible domains of emergent properties, of which mental properties are the leading candidates.").

81. *See, e.g.*, Morse, *Common Sense*, *supra* note 23, at 48; Morse, *Inevitable Mind*, *supra* note 2, at 34 (though the "mind/brain … is only one substance, it has both physical and mental properties [and the] latter are emergent and cannot be reduced fully to the former").

82. *See supra* note 78.

intention to commit a crime. The way that the redness of cherries "looks" or the intention to commit a crime subjectively "feels" is a phenomenological experience that *emerges* from the physical firings of neurons but which is in itself non-physical and is therefore different and distinct from ("over and above") the neuron firings that underlie it.[83] Similarly, other patterns of synaptic discharges within populations of neurons can give rise to other emergent mental-state experiences (in addition to seeing, feeling and intending), such as recalling, imagining, reasoning, believing and all the other furniture of our mental lives. None of this phenomenological richesse is considered to be inconsistent with physicalism because all of these emergent mental states, although not themselves physical stuff located in time and space, can nonetheless be said to be properties (albeit *non-physical* properties) of the underlying, physical neuronal activity.

While property dualism and emergentism may be reasonable ways to conceptualize certain aspects of our lived experience, there are problems with using them to reconcile physicalism with *autonomous* mental causation. The reason is that one of the "rules" of property dualism is that the *dependent* (or emergent) properties have to co-vary with the independent ones.[84] For example, two paintings cannot be physically identical in every respect and still depict two different scenes or have two different sets of aesthetic qualities.[85] Accordingly, property dualism does not authorize the conclusion that a given pattern of neuronal firings can give rise to two or more different mental states, or that sequences of mental states can wander autonomously of their physical base. A given pattern of neuronal firings cannot spawn different mental states like a frog spawns tadpoles, which are free to swim off wherever they please. If mental states are rooted in brain states, as property dualism contends, they must co-vary with the brain states that they are properties of—and sequences of mental states are therefore subject to the same computational rules as their underlying brain states. Any physical causation by such mental states would, accordingly, not be autonomous. On the contrary, even if such mental states are causally efficacious, they would be mere conduits carrying behavioral instructions generated and determined by the computational brain. The behavior that is

83. *See* KIM, PHYSICALISM, *supra* note 10, at 154.

84. *See* DAVID LEWIS, ON THE PLURALITY OF WORLDS 14 (1986) ("[N]o two [dot-matrix] pictures could differ in their global properties without differing, somewhere, in whether there is or there isn't a dot."); KIM, PHYSICALISM, *supra* note 10, at 20.

85. KIM, PHYSICALISM, *supra* note 10, at 20.

caused by such mental states would still be determined (though at one remove) by the underlying brain activity.

Likewise, emergentism (a form of property dualism) does not mean that mental states can emerge from physical brain states like a genie from a bottle, wielding magical powers to override physical laws.[86] There is sometimes confusion in this regard owing to the fact that there seem to be "laws" or lawlike regularities at emergent levels (for example, "laws" of biology, psychology, or aesthetics) that are not reducible to the physical laws that apply at underlying levels. For example, how can physical laws ever explain why one painting of Paris is beautiful and another is atrocious? Obviously, we can cognitively discern "higher-level" laws as we observe that some emergent entities fit together like a lock and key or otherwise interact with regularity while others do not, depending on the properties and configurations of the matter that comprises them. But there is no evidence that these lawlike regularities of interactions that are cognitively discerned among emergent entities, such as individual biological beings, are ever able to override the physical laws that are applicable at the base level. There is no evidence, in other words, that distinct "higher-level" forces ever emerge with ability to move or displace a single molecule—much less an arm or a leg—in opposition to the ordinary physical forces described in physical laws.

In sum, both property dualism and emergentism take the view that physicalism does not rule out the possibility that physical states and objects can have non-physical properties and that mental states are non-physical properties of the neuronal activity that gives rise to them. As such, however, mental states are dependent on the physical substrate that they are properties of, not autonomous from it. Neither property dualism nor emergentism is, therefore, able to reconcile the contradiction between mental-causation autonomy and physicalism.[87] That is, neither is able to change the fact that

86. "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin, when Aladdin rubbed his lamp." THOMAS HUXLEY, LESSONS ON ELEMENTARY PHYSIOLOGY 193 (1866).

87. There is, of course, much more to be said about emergence, *see, e.g.*, Timothy O'Connor & Hong Yu Wong, *Emergent Properties, The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed., Summer 2015), https://plato.stanford.edu/archives/sum2015/entries/properties-emergent/. Whatever else may be said about it, however, this much seems clear: The individuation of objects and events that we see in nature is imposed by us as observers and not an ontological given. Emergence *appears* to occur essentially because of the way we conceptualize, individuate and give names to higher-level abstractions, regularities and aggregations that we cognitively discern in the weltering, undifferentiated chaos of nature. There is, however, nothing in our conceptualizing, individuation and name-giving per se that would allow what we see as a process in nature to actually jump the rails of fundamental physical law and ramble off to establish a law unto itself. Emergence is not, in other words,

a commitment to physicalism excludes mental-state autonomy and, accordingly, excludes the possibility that mental causation (even if it occurs) could provide moral significance to crimes and other physically-determined bodily acts.

III. "NONREDUCTIVE" PHYSICALISM

Another strategy to escape the contradiction at the core of the neuroskeptical thesis has been to appeal to a custom-tailored alternative to actual physicalism known as *nonreductive* physicalism.[88] Nonreductive physicalism accepts that "we inhabit a thoroughly material, physical universe,"[89] but it contends nonetheless that, though the "mind/brain … is only one substance, it has both physical and mental properties [and the] latter are emergent and cannot be *reduced* fully to the former."[90] The nonreductive physicalist believes, in other words, that it is a mistake to think that intentions, reasons and other mental states are "nothing more than" electrical activity in the neurons of the brain. Rather, these mental states are something "over and above" the underlying brain processes that bring them about.[91] It is accordingly not possible to fully explain or understand mental states in terms of the brain states that co-occur with them—unlike, say, the workings of a laptop computer, which can be fully understood and explained in terms of the transistors and other electronics that comprise it. The idea is that there is more to mental states than just an aggregation of physical activity in the brain.[92]

---

the magic key that unlocks the conundrum that faces neuroskeptics as a result of their contradictory stances toward the physicalism of human behavior.

88. *See generally* KIM, PHYSICALISM, *supra* note 10, at 94-98. Professor Morse, for example, states that he is "most attracted" to nonreductive physicalism. Morse, *Common Sense*, *supra* note 23, at 48. Though treated separately in the text, property dualism, emergentism and nonreductive physicalism are not distinct hypotheses but are ways of looking at the same hypothesis, namely, that property dualism is a feature of reality because non-physical properties can emerge from physical arrangements and systems and these emergent non-physical properties are not necessarily reducible to underlying physical properties.

89. Morse, *Common Sense*, *supra* note 23, at 47; Morse, *Inevitable Mind*, *supra* note 2, at 33.

90. Morse, *Common Sense*, *supra* note 23, at 48; Morse, *Inevitable Mind*, *supra* note 2, at 34 (emphasis added). "There is no consensus on exactly how nonreductive physicalism is to be formulated, for the simple reason that there is no consensus about either how physicalism is to be formulated or how we should understand reduction." KIM, PHYSICALISM, *supra* note 10, at 33, 94, 98. It seems, however, clear that nonreductive physicalism means, at the very least, that mental states are in a separate ontological class from physical reality and that they do not consist solely of physical matter, arrangements of physical matter or physical states.

91. *See* KIM, PHYSICALISM, *supra* note 10, at 34.

92. To draw the key contrast between ordinary physicalism and nonreductive physicalism: In ordinary physicalism, mental events must co-vary in every particular with the underlying physical events on which they depend and therefore be reducible to them, and not something "over and above" them.

To be sure, nonreductive physicalism in itself is not necessarily an unsound hypothesis and, indeed, it is "probably the dominant view among specialists."[93] And there is good reason. After all, though we do not know the precise quiddity of consciousness, few would seriously deny that people have minds in addition to brains[94] or that the way a person experiences conscious mental events is something "over and above" the physical activity that goes on in the neurons. Mental states as subjectively experienced have a non-physical nature that no one yet has managed to describe in purely physical terms. For example, it seems distinctly unlikely that what it is like to be conscious of the redness of cherries, the melodies of songbirds, or the flavor of pinot noir is the same identical thing as the physical brain activity (synaptic firing) that gives rise to these sensations. It has never been persuasively shown, at any rate, *how* conscious sensations such as these can be "reduced" to or fully described in terms of purely physical events. Yet, the existence of such sensations cannot be credibly denied, and our consciousness of them is, when you get down to it, the only thing our minds can actually directly know.[95] They are, in a sense, our entire lived experience. The apparently universal existence of conscious awareness in human experience is reason enough in itself to make nonreductive physicalism the "dominant view."[96]

See KIM, *supra* note 1, at 212-16; KIM, PHYSICALISM, *supra* note 10*,* at 33-34. Accordingly, the occurrence, content and flow of mental states would, on analysis, always turn out to mirror underlying physical events because (according to physicalism) everything having a concrete existence can be fully explained and accounted for in physical terms and must, therefore, have a physical cause (or no cause). *See id.* One important consequence of this feature of physicalism is that, if physicalism is true, the mind could not wander off on paths of its own and autonomously cause behavior that the physical operations of the brain would not cause anyway. Nonreductive physical posits that mental states can have non-physical properties whose autonomy is not similarly constrained.

93. Morse, *Inevitable Mind*, *supra* note 2, at 34; *see also* Morse, *Common Sense*, *supra* note 23, at 47.

94. *See* Galen Strawson, *The Consciousness Deniers*, N.Y. REV. BOOKS, Mar. 13, 2018, http://www.nybooks.com/daily/2018/03/13/the-consciousness-deniers/ [https://perma.cc/JP93-PNXR] ("What is the silliest claim ever made? The competition is fierce, but I think the answer is easy. Some people have denied the existence of consciousness.").

95. *See, e.g.*, Arthur Schopenhauer, II WORLD AS WILL AND REPRESENTATION 5 (E.F.J. Trans. 1968) (1859). As physicist Arthur Eddington once wrote: "But no one can deny that mind is the first and most direct thing in our experience, and all else is remote inference." ARTHUR EDDINGTON, THE NATURE OF THE PHYSICAL WORLD 281 (Cambridge Univ. Press 1928), http://henry.pha.jhu.edu/Eddington.2008.pdf [https://perma.cc/U86H-C6W3].

96. Just to head off a possible accusation of inconsistency, let me hasten to clarify that there is an important difference between relying on introspection as evidence that non-physical mental states exist and relying on introspection as evidence that they are causally efficacious. In the first case, we are saying that conscious sensations that we experience as direct *apercus* are, as sensations, proof of their own existence. But mental causation is not a direct *apercu* but, at best, only an inference that is drawn from *apercus*, and a fallacious one at that. *See Mental Causation Fallacy, supra* note 3, at 224-38.

It is a bit of a reach, however, to say that the nonreducibility of mental states supports the mental causation hypothesis. Merely accepting the truth of nonreductive physicalism does not resolve the inherent contradiction facing the neuroskeptic who wants to claim the mantle of physicalism but still suppose that mental causation can occur autonomously of the physical (and, therefore, be morally significant).[97] After all, nonreductive physicalism does not in itself imply mental causation or, for that matter, any causation at all, either from mind to body or from body to mind. It holds only that mental states can have properties that are not reducible to physical properties of co-occurring brain states; it says nothing about the causal powers of those properties. It is one thing to observe that people have conscious awareness and infer from it that non-physical (mental) phenomena must therefore be able to emerge out of physical (neuronal) events with properties beyond those inhering in the physical. But it is pure non-sequitur to turn that emergence into a causal arrow in the opposite direction and infer that non-physical mental states can produce physical events (such as bodily movements) on their own, without depending on underlying brain events to do the actual causal work.[98] The latter inference is like noting that electrical activity in your television determines what appears on the screen and then concluding that the events you see on the screen can change what happens inside the set.[99]

The inference that conscious awareness cannot be fully reduced to co-occurring brain activity is based on what we know from neuroscience together with the vivid experience of consciousness that it would be quixotic to deny. By contrast, the inference that non-physical mental-state properties can produce physical bodily movements is pure speculation, with no basis in evidence other than bare correlation.[100] To infer mind-to-body causation from this correlation alone, absent any indication of causal mechanism, is to reason *post hoc ergo propter hoc*—a logical fallacy.[101] Indeed, it is worse than that: In order for intentions, reasons and other nonreducible mental

97. As previously discussed, the causal autonomy of mental states is important because mental causation in the causal chain of behavior would seem to have no special moral significance if the causal effects of mental events exactly duplicated the effects that would be produced by physical processes anyway. *See supra* notes 53-76 and accompanying text.

98. *See, e.g.*, JOHN R. SEARLE, MIND: A BRIEF INTRODUCTION 114, 209 (2004) ("[M]ental processes…have no causal powers in addition to those of the underlying neurobiology" and "there is no causal efficacy to consciousness that is not reducible to the causal efficacy of its neuronal basis.").

99. Of course, many computers now have "touchscreens" that are, in effect, transparent input devices placed *on top of* the screen itself. As input devices, these can of course affect what happens inside the computer. But the images on the actual screen can *not* affect what happens inside.

100. *See Mental Causation Fallacy*, *supra* note 3, at 224-38.

101. *See Mental Causation Fallacy*, *supra* note 3, at 224-34 and accompanying text.

properties to produce bodily movements *autonomously*—i.e., movements that the brain's physical processes do not already direct the production of—the mental properties would have to be able to cause physical events to occur in *opposition* to the physical brain's directions and hence in opposition to physical law. Nonreductive physicalism does not resolve but, if anything, exacerbates the contradiction between the mental-causation hypothesis and the commitment to physicalism that predominates in modern thought.

So, while nonreductive physicalism may be an attractive hypothesis for reconciling a physicalist view of the world with the subjective experience of consciousness, which science has not explained, it does not support a leap to the conclusion that nonreducible mental states can cause bodily movement that science *can* explain. Accordingly, even if nonreductive physicalism is true, as it probably is (in limited applications), it does not support a valid inference of mental causation. The case for mental causation can only stand on the premise that physical events (like human conduct) can be produced by non-physical causes. But this is a premise that is not only at odds with physicalism but, even more importantly, remains utterly undemonstrated in any context anywhere in the Universe.

IV. WHY THE NEUROSKEPTIC'S CONTRADICTION IS IMPORTANT

The friends of mental causation may claim to accept the truth of physicalism, but their embrace of it is an uneasy one. Why should this matter? After all, although the truth of physicalism is strongly supported by observations from throughout the Universe, it is not an ontological certainty,[102] and it is hard to see how *any* physicalist position can be unqualified as long as the quiddity of consciousness remains unknown. Indeed, for most of human history nearly everyone believed that physicalism was false. The existence of non-physical forces and paranormal powers were taken for granted, and stories about them abounded. The bewildering mosaic of contradictory beliefs that people still firmly hold today, religious and non-religious, are testimony to the rich spiritual life, capacities and imagination of humankind.

To be sure, everyone has a right to an abiding faith in the reality of non-physical forces including autonomous mental causation, and those who hold such beliefs are not to be disdained. But the question in the context of criminal justice is not about freedom of conscience nor is it just an academic matter of metaphysical debate. The question is, bluntly, whether the legal

102. *See* Papineau, *supra* note 33.

system of a secular government can justify the systematic infliction of hardship, deprivation and misery on its people based on world-views that do not include a strong presumption of physicalism and a definite preference for physical (versus non-physical) explanations of the events that occur.[103] While all are entitled to their beliefs, we should hesitate to accept moral instruction from those who reject such a common-sense presumption and whose commitment to basic physical reality is therefore so open to question. It is at least prudent to view with suspicion their assertion that people deserve to be made to suffer because their actions can be produced by their intentions and reasons rather than ordinary physical causes. It is particularly prudent if one does not share the faith-system and shaky commitment to physicalism on which that position is based.[104]

---

103. *See* KIM, PHYSICALISM, *supra* note 10, at 156 ("Motivations for introducing entities other than material things vary—from supposed philosophical requirements in connection with certain issues, for example, the persistence of persons over time, the possible survival of bodily death, and they special directness of knowledge of one's own mind, to religious imperatives and mystical intimations."). Oddly, Kim does not mention the impetus to satisfy the retributive appetite for bringing suffering and misery to human beings in the guise of justice, since this seems to be (though often cloaked in high sounding words like responsibility) a primary motivating factor, especially among authors who discuss the topic from a law-related perspective.

104. Just to be clear (because it always comes up): There is nothing in the findings of neuroscience to suggest that confinement and other coercive measures are not sometimes necessary in the interest of protecting the public from persons who present an unreasonable or socially intolerable risk of harm to others. *See Mental Causation Fallacy*, *supra* note 3, at 243-53. But, as further discussed *id.*, the nature and quality of those measures may be very different if they are treated as regrettable necessities rather than as deserved.