ALABAMALAW
THE UNIVERSITY OF ALABAMA®

Alabama Law Scholarly Commons

Working Papers | Faculty Scholarship

3-14-2011

Neuroscientific Challenges to Retributivism

Michael S. Pardo
University of Alabama - School of Law, mpardo@law.ua.edu

Follow this and additional works at: https://scholarship.law.ua.edu/fac_working_papers

Recommended Citation

Michael S. Pardo, *Neuroscientific Challenges to Retributivism*, (2011).
Available at: https://scholarship.law.ua.edu/fac_working_papers/290

This Working Paper is brought to you for free and open access by the Faculty Scholarship at Alabama Law Scholarly Commons. It has been accepted for inclusion in Working Papers by an authorized administrator of Alabama Law Scholarly Commons.

THE UNIVERSITY OF
ALABAMA
SCHOOL OF LAW

Neuroscientific Challenges to Retributivism

Michael S. Pardo
Dennis Patterson

THE FUTURE OF PUNISHMENT (Thomas Nadelhoffer ed., Oxford University Press, forthcoming)

This paper can be downloaded without charge from the Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract=1783823

Neuroscientific Challenges to Retributivism

Michael S. Pardo*

Dennis Patterson**

Theories of criminal punishment provide accounts that purport to legitimate and justify criminal punishment. The theoretical project takes as its impetus that criminal punishment is a form of state-sponsored violence; under the auspices of criminal punishment, states inflict pain and suffering on citizens—depriving them of life, liberty, or property to which they would otherwise be entitled—for their transgressions of the criminal law. What would make such coercive actions by the state legitimate? When is the exercise of this power justified? Finally, how much punishment is justified in particular circumstances? Much ink has been spilled trying to answer these difficult questions. Although a variety of different theories and approaches have been proposed throughout the ages, modern criminal-law theory centers around two groups of punishment theories.[1] Broadly construed, theories in the first group purport to justify

* Henry Upson Sims Professor of Law, University of Alabama School of Law. Prof. Pardo thanks Dean Ken Randall and the University of Alabama Law School Foundation for generous research support.

** Professor of Law and Chair in Legal Philosophy and Legal Theory, European University Institute, Florence; Board of Governors Professor of Law and Philosophy, Rutgers University, New Jersey, USA; Professor of Law and Chair in International Trade and Legal Philosophy, Swansea University, Wales, UK.

[1] In an insightful and clarifying recent article, Mitchell Berman observes that despite the

> rich diversity of justificatory theories, including deterrence (Bentham and Beccaria), reform (Plato), retribution (Kant), annulment (Hegel), and denunciation (Durkheim[, a] striking feature of twentieth century punishment theory . . . has been the steady and generally successful pressure to fold this seeming multiplicity of justifications into a simple dichotomy of justifications that at least appears to mirror the fundamental organizing distinction in moral theory between consequentialism and deontology.

Mitchell Berman, Two Types of Retributivism, in The Philosophical Foundations of the Criminal Law 3 (Duff & Green, eds, forthcoming), available at: http://ssrn.com/abstract=1592546.

Electronic copy available at: http://ssrn.com/abstract=1783823

punishment on "consequentialist" grounds. Despite important differences among them, theories in this group rely on the beneficial social consequences that are claimed to flow from criminal punishment (primarily, reduced future crime) because of its deterrent, incapacitating, or rehabilitating effects.[2] Broadly construed, theories in the second group purport to justify punishment on "retributivist" grounds. Theories in this group, despite important differences among them, rely on the notion that criminal offenders somehow *deserve* punishment proportionate with their transgressions of the criminal law.[3]

Although debates rage on within each side and across the consequentialist-retributivist divide, most theorists acknowledge some role for both types of considerations. The United States Supreme Court has also explained that, as a matter of constitutional law, the federal and state governments may as a general matter rely on both types of considerations.[4] Empirical work suggests that subjects support both rationales for punishment (although particular punishment decisions may be more consistent with retributivist rationales).[5] We do not wish to take sides or offer new arguments in favor of retributivist or consequentialist theories of criminal punishment.

---

[2] The classic deterrence-based accounts are Cesare Baccaria, On Crimes and Punishments (1764) and Jeremy Bentham, An Introduction to the Principles of Morals and Legislation (1789). For an overview of more recent consequentialist rationales, see Anthony Duff, Legal Punishment, Stanford Encyclopedia of Philosophy (2008), available at: http://plato.stanford.edu/entries/legal-punishment/ .

[3] The classic retributive account is Kant's. See Immanuel Kant, The Metaphysics of Morals 1797 (Gregor trans. 1996). See also Berman, supra note 1, at 6 (referring to the "desert claim" as the "core retributivist contention" that "punishment is justified by the offender's ill-desert.") For overviews, see Duff, supra note 2; David Wood, Punishment: Nonconsequentialism, 5/6 Philosophy Compass (2010), available at http://ssrn.com/abstract=1659453.

[4] See Harmelin v. Michigan, 501 U.S. 957, 1001 (Kennedy, J., concurring) ("the Eight Amendment does not mandate adoption of any one penological theory . . . The federal and state criminal systems have accorded different weights at different times to penological goals of retribution, deterrence, incapacitation, and rehabilitation.")

[5] See, e.g., Kevin M. Carlsmith, John M. Darley & Paul H. Robinson, Why Do We Punish? Deterrence and Just Desserts as Motives for Punishment, 83 Journal of Personality and Social Psychology 284, 294 (2002) (noting that "[w]hen asked about just deserts and deterrence, participants generally supported both perspectives" and "[p]eople seemed to support these two philosophies and generally to have a positive attitude toward both").

Electronic copy available at: http://ssrn.com/abstract=1783823

Rather, we examine the relationship between neuroscience and this theoretical project, focusing in particular on arguments and inferences regarding criminal punishment drawn from current neuroscientific data.

Recently, claims have been made that neuroscientific evidence may indeed resolve—or at least make a significant contribution to—this debate. In a series of articles, Joshua Greene and colleagues have argued that neuroscientific evidence challenges and ultimately undermines (or will undermine) retributivist theories of punishment.[6] Two separate arguments are made for this anti-retributivist conclusion. The first argument appeals to neuroscientific data about the brain activity of those making punishment decisions. According to this argument, "retributive" punishment decisions are correlated with brain activity associated with more "emotional" rather than "cognitive" processes, and, therefore, this somehow undermines their status.[7] The structure of this argument is similar to ones also made regarding moral and economic decision-making: namely, one theoretical "type" of decision is correlated with a "type" of brain activity or process and is thus impugned because of that correlation.[8] In these other contexts, non-utilitarian (deontological or otherwise) judgments are also impugned because of their correlations with

---

[6] Joshua D. Greene, The Secret Joke of Kant's Soul, in Moral Psychology, Vol 3: Emotion, Disease, and Development (Sinnott-Armstrong ed., 2007); Joshua Greene & Jonathan Cohen, For Law, Neuroscience Changes Nothing and Everything, in Law and the Brain (Seki & Goodenough eds, 2006); see also Azim F. Shariff, Joshua D. Greene & Jonathan W. Schooler, His Brain Made Him Do It: Encouraging a Mechanistic Worldview Reduces Punishment (forthcoming, include cite).

[7] See Greene, Secret Joke, supra note 6, at 50-55.

[8] See, e.g., Joshua D. Greene, The Neural Bases of Cognitive Conflict and Control in Moral Judgment, 44 Neuron 389 (2004); Joshua D. Greene, An fMRI Investigation of Emotional Engagement in Moral Judgment, 293 Science 2105 (2001); Alan Sanfey et al., The Neural Basis of Economic Decision-Making in the Ultimatum Game, 300 Science 1755 (2003). For critiques of attempts to draw normative conclusions from such studies, see Richard Dean, Does Neuroscience Undermine Deontological Theory? 3 Neuroethics 43 (2010); Michael S. Pardo & Dennis Patterson, Philosophical Foundations of Law and Neuroscience, 2010 U. Illinois L. Rev. 1211 (2010); Selim Berker, The Normative Insignificance of Neuroscience, 37 Phil. & Public Affairs 293 (2009); F.M. Kamm, Neuroscience and Moral Reasoning: A Note on Recent Research, 37 Phil. & Public Affairs 330 (2009).

"emotional" brain activity.[9] Although there are important differences between the criminal-punishment context and the moral and economic contexts—differences we explore below—the argument in the punishment context fails for similar reasons. First, the success or failure of retributivism does not depend on the success or failure of any particular moral theories, and thus *even if* neuroscience undermined deontological moral theories it would not necessarily also undermine retributivism. Second, retributivism does not depend on which areas of the brain are associated with punishment decisions. Brain activity does not provide criteria for whether punishment decisions are correct or just, nor does the fact that retributivist decisions are associated with "emotional" decision-making provide evidence that the decisions are incorrect or unjust.

A second challenge to retributivism appeals to neuroscientific data to undermine retributivist intuitions *indirectly* by undermining *directly* the "free will" intuitions on which, it is claimed, retributivist theories depend.[10] This argument requires some unpacking before we can evaluate its underlying assumptions. Before we explore the details of this argument, however, it is important to understand more generally how the argument relates to current criminal-law doctrine. The criminal law presupposes "folk psychological" explanations of human action: neuroscientific data may provide inductive empirical evidence that is relevant for deciding issues within that conceptual framework.[11] A variety of proposals have suggested ways in which neuroscience may provide evidence for deciding issues within this current conceptual

---

[9] See supra note 8.

[10] Greene & Cohen, For Law, Neuroscience Changes Nothing and Everything, supra note 6.

[11] See Stephen J. Morse, Criminal Responsibility and the Disappearing Person, 28 Cardozo Law Review 2545, 2253-54 (2007) ("The law's view of the person is thus the so-called 'folk psychological' model: a conscious (and potentially self-conscious) creature capable of practical reason, an agent who forms and acts on intentions that are the product of the person's desires and beliefs. We are the sort of creatures that can act for and respond to reasons.").

framework, including issues regarding mens rea, insanity, competence, voluntariness, and lie detection. Importantly, however, the inferences and conclusions drawn from the neuroscientific data must not run afoul of the conceptual contours of that framework in order to contribute meaningfully to these doctrinal issues.[12]

At the theoretical level of criminal punishment, however, neuroscience may offer a deeper and more radical challenge to the entire doctrinal framework of the criminal law by undermining the "folk psychological" assumptions on which it is based and, with it, theories of punishment that depend on such assumptions. This is the nature of the second challenge to retributivism. It concedes that the current doctrinal edifice of the criminal law remains largely unshaken by neuroscience. It argues, however, that important aspects of the doctrinal edifice depend upon a retributivist foundation—which in turns rests upon a non-deterministic, free-will foundation—and thus, so the argument goes, if neuroscience can bring down the non-deterministic, free-will foundation it will also bring down retributivism and the legal doctrine built upon it. We discuss several problems with this argument and argue that if neuroscience has the potential to cause the changes Greene and Cohen predict, it will do so by fostering a number of unwarranted and problematic inferences and ought to be resisted.

In this chapter, we first discuss theories of criminal punishment in more detail and then evaluate each of the two arguments for the purported neuroscientific overthrow of retributivism.

I. A Brief Taxonomy of Theories of Criminal Punishment

Theories of criminal punishment are primarily normative accounts that purport to answer the question of when the state is justified in subjecting citizens (and non-citizens) to criminal punishment, and derivatively, how much punishment is justified when punishment in general is

[12] For a discussion, see Pardo & Patterson, supra note 8.

warranted.[13] In addition to this normative project, theories of punishment may also be directed at answering explanatory questions as to why the state would engage in acts of criminal punishment—whether it is justified or not—and why it would choose particular forms and amounts of punishment.[14] There is also a distinct conceptual project of delineating the scope of what constitutes punishment.[15]

One dominant strategy for answering such questions focuses on the future consequences of punishment. Under this "forward looking" strategy, the perceived beneficial consequences include deterring others not punished from committing similar acts in the future, and preventing (or reducing the likelihood) that those punished will commit future crimes by rehabilitating, incapacitating, or deterring them specifically. Under this consequentialist strategy, punishment is not an end in itself but serves an instrumental value in bringing about the social good of reducing future crime, and punishment may be justified to the extent the positive social benefits it brings about exceed the harms it causes.[16] Moreover, specifics about whom should be punished, how, and how much, may be justified under this strategy based on whatever would bring about the socially optimal or desirable level of benefits to costs. In addition to these normative issues, a consequentialist account may also explain why the state would choose to exercise its right to punish (assuming it is justified in doing so) and would undergo the expenses of doing so (including expenses to those punished and citizens generally).[17]

---

[13] See Duff, supra note 2; Berman, supra note 1.

[14] See John Bronsteen, Retribution's Role, 84 Indiana L.J. 1129 (2009).

[15] See H.L.A. Hart, Punishment and Responsibility (1968).

[16] See Berman, supra note 1.

[17] See Bronsteen, supra note 14.

The second dominant strategy for answering such questions focuses on the acts (together with the mental states and surrounding circumstances) of those subjected to criminal punishment. Under this "backward looking" strategy, actions by those who violate the dictates of the criminal law are such that the actor may deserve punishment and ought to punished (and conversely those who do not violate the criminal law do not deserve punishment and should not be punished), regardless of whether any future beneficial consequences will follow.[18] Under this strategy, it is the fact that the guilty defendant deserves criminal punishment that justifies the state's actions.[19] Retributivists may cash out exactly how desert justifies punishment in a variety of ways. For instance, punishment of those who deserve it may have some innate, intrinsic worth.[20] The desert aspect may also serve a particular type of instrumental value that justifies punishment—for example, it may serve to "cancel out," denounce the criminal acts, or express the communities disapproval for the various acts.[21] Or under a more "pure" form of retributivism, the desert aspect may justify punishment regardless of whether punishment itself has any intrinsic worth or serves any other intrinsic value.[22] Turning to the specifics of punishment, the retributivist strategy purports to justify whom should be punished, how, and how much by appealing to whether the person is in fact guilty and, if so, the amount of punishment

[18] See Berman, supra note 1.

[19] Id.

[20] See Michael Moore, Placing Blame (2007).

[21] Berman refers to justifications that depend on intrinsic worth or these other goals as "instrumental retributivism." See supra note 1, at 9.

[22] Id. at 16-19.

that is proportional to their culpability or ill-desert.[23] In addition to these normative issues, a retributivist account may also explain why the state chooses to punish and chooses to do so in the ways that it does. Under this explanatory account, such punishment tracks the intuitions of citizens about what is just and may also reduce acts of vengeance and reciprocal violence.[24]

The initial distinction between these two strategies raises a number of further issues. First, the considerations at issue under each strategy may play a variety of theoretical roles. They may each be taken to provide one consideration in whether punishment is justified in a particular context.[25] Under such a view, the criteria of any particular strategy may be neither necessary nor sufficient to justify punishment. Or each strategy may be taken to provide a *necessary* condition for justifying punishment.[26] Or each may be taken to provide a *sufficient* condition for justifying punishment.[27] Second, the strategies may also combine and interact in various ways. For example, each may provide a "constraint" on the other that would defeat otherwise legitimate punishment—punishment that would produce good consequences, all things considered, may be illegitimate if it punishes someone more than they deserve, and punishing someone as much as they deserve may be illegitimate if it would otherwise lead to terrible social consequences.[28]

---

[23] Under retributivist theories, "proportionality" may be cashed out in various ways. See Alice Ristroph, Proportionality as a Principle of Limited Government, 55 Duke L.J. 263, 279-84 (2005).

[24] See Paul H. Robinson & John M. Darley, Intuitions of Justice: Implication for Criminal Law and Justice Policy, 81 S. Cal. L. Rev. (2007).

[25] Under this view, the strategies are consistent with each other.

[26] For example, satisfying the criterion of desert may be required to justify punishment, but it alone may not be sufficient.

[27] For example, deterrence may provide a sufficient condition for punishment under some conceptions, but it may not be necessary.

[28] Under such conceptions, each strategy may provide "defeasible" conditions for justifying punishment. For example, desert might be taken to justify punishment unless it can be shown that the punishment will lead to more crime. For further details on different possible ways to conceptualize retributivism, see Berman, supra note 1; Larry

Finally, although the two strategies lend themselves to the familiar distinction in moral theory between utilitarian and deontological theories, they are conceptually distinct.[29] One may believe that deontological considerations ground moral theory or make particular moral judgments right or wrong (true or false)[30] and also think (consistently) that the state is not justified in punishing for retributivist reasons.[31] Similarly, one may believe that utilitarian considerations ground moral theory or particular moral judgments and also think (consistently) that the state is justified is engaging in criminal punishment for retributivist reasons.[32]

As our language is meant to suggest, we do not intend to take sides in these debates or to argue that any particular theories in these categories succeed or fail on their own terms. Our aim in this brief section has been simply to explicate the theoretical issues sufficiently in order to properly assess the arguments for how neuroscientific data affects the theoretical issues. We now turn to these arguments.

II. Brains and Punishment Decisions

Alexander & Kimberly Kessler Ferzan, with Stephen Morse, Crime and Culpability: A Theory of Criminal Law (2009) (distinguishing "mild," "moderate," and "strong retibutivism"); Kenneth W. Simons, Retributivism Refined—or Run Amok?, 77 U. Chi. L. Rev. (2010) (book review essay); Michael T. Cahill, Punishment Pluralism, in Retributivism: Essays on Theory and Policy (White ed., forthcoming), available at http://papers.ssrn.com/sol3/abstract=1705682.

[29] Larry Alexander & Michael Moore, Deontological Ethics, Stanford Encyclopedia of Philosophy (2007), available at http://plato.stanford.edu/entries/ethics-deontological/ ; Berman, supra note 1, at 4-5.

[30] This point holds regardless of whether moral truths are understood in realist or anti-realist terms.

[31] See Alexander & Moore, supra note 29 ("Retributivism has two aspects: (1) it requires that the innocent not be punished, and (2) it requires that the guilty be punished. One could be a deontologist generally and yet deny that morality has either of these requirements.").

[32] This may be one normative implication of Paul Robinson's work on "empirical desert." See Paul H. Robinson, Empirical Desert, in Criminal Law Conversations 29 (Robinson, Garvey & Ferzan eds., 2009). See also Alexander & Moore, supra note 29 ("a retributivist might alternatively cast these two states of affairs (the guilty getting punished and the innocent not getting punished) as two intrinsic goods, to be traded off both against each other (as in burden of proof allocation) and against other values. Some retributivists urge the latter as a kind of explicitly 'consequentialist retributivism.'").

The relationship between the psychology of punishment decisions and the normative project of justifying punishment is a complicated one. Understanding how or why punishment decisions are made does not necessarily tell us how such decisions ought to be made or whether they are justified. A further argument is needed about *how* empirical information is supposed to bear on the normative, theoretical questions. On the one hand, the empirical evidence might be thought to provide positive support for particular punishment decisions or general theories—or at least constrain possibilities—by illustrating the way most people would decide to punish or what most people would judge to be fair or just regarding punishment decisions.[33] Such a project might appeal to actual, hypothetical, or idealized punishment situations or conditions. On the other hand, the empirical evidence might be thought to undermine particular decisions or general theories if the evidence shows that the decisions made (or implied by a general theory) are produced by an "unreliable" or otherwise defective process.[34]

The relationship between neuroscience and the normative questions regarding criminal punishment is more complicated still. The theoretical move from neurological processes to normative conclusions about criminal punishment requires an argument not only from what people are doing when they decide to punish to whether they are justified in doing so; it requires an argument from what their brains are doing, to what they are doing, to whether what they are doing is justified. One way to bridge these conceptual gaps is the one proposed by Joshua Greene and colleagues with regard to moral decision-making. Under this framework, decisions

---

[33] See Robinson, supra note 32.

[34] Selim Berker suggests that the "best-case scenario" for undermining certain moral intuition based on neuroscience would be to show that brain areas correlated with the intuitions are also correlated "obvious, egregious error[s] in mathematical or logical reasoning," but he concludes that even this claim depends on further assumptions and philosophical argument. See Berker, supra note 8, at 329.

are made through one of two psychological processes: a "cognitive" one and an "emotional" one.[35] Each of these processes is associated with different patterns of brain activity or areas in the brain; fMRI data is used determined which brains areas appear to be more activated during particular decisions and, based on this data, inferences are drawn about which psychological process was used to make the decision.[36] An additional argument is then needed to move from the implicated psychological process to normative conclusions about criminal punishment.[37]

As with psychological evidence, neuroscientific evidence regarding punishment decisions may be used to bolster or to challenge claims or theories about punishment. Examining a wide swath of both psychological and neuroscientific studies (including his own), Joshua Greene argues that neuroscience challenges retributivist theories of punishment (and supports consequentialist theories).[38] Greene defines the two approaches to justifying punishment broadly, with consequentialists asserting that punishment is "justified solely by its future

---

[35] Greene, Secret Joke, supra note 6, at 40. The "cognitive" process involves (1) "inherently neutral representations, ones that do not automatically trigger particular behavioral responses or dispositions"; is (2) "important for reasoning, planning, manipulating information in working memory, impulse control"; and is (3) "associated with . . . the dorsolateral surfaces of the prefrontal cortex and parietal lobes." Id. By contrast, the "emotional" process (1) triggers automatic responses and dispositions, or are "behaviorally valenced"; (2) is "quick and automatic"; and (3) is "associated with . . . the amygdala and the medial surfaces of the frontal and parietal lobes." Id. at 40-41.

[36] The data are generated from a number of experiments in which subjects were presented with a series of vignettes involving moral dilemmas (e.g. variations on the "trolley problem") and in which the "consequential" and "deontological" answers appear to diverge. See supra note 8. Subjects had their brains scanned while deciding the cases, and the answers to the vignettes were compared with the brain activity of subjects. Greene et al. made two predictions, which were to a large extent borne out by the data: (1) "consequentialist" judgments would be correlated with "cognitive" brain activity and "deontological" judgments would be correlated with "emotional" brain activity, and (2) "consequentialist" judgments would on average take longer to make than "deontological" ones. Our analysis does not take issue with either of these conclusions; we grant them for the sake of our arguments. But for discussion of some potential methodological and empirical issues raised by the studies, see Berker, supra note 8 and John Mikhail, Moral Cognition and Computational Theory, in Moral Psychology, Vol 3: Emotion, Disease, and Development (Sinnott-Armstrong ed., 2007).

[37] Similarly, an additional argument is needed to move from the data and psychological processes to normative conclusions in morality or economics.

[38] Greene, Secret Joke, supra note 6.

beneficial effects," and retributivists asserting that its "primary justification" is "to give wrongdoers what they deserve based on what they have done, regardless of whether such distribution will prevent future wrongdoing."[39] He next examines "the psychology of the criminal punisher," and summarizes as follows:

> People endorse both consequentialist and retributivist justifications for punishment in the abstract, but in practice, or when faced with more concrete hypothetical choices, people's motives appear to be emotionally driven. People punish in proportion to the extent that transgressions make them angry.[40]

Assuming it is true that people's punishment decisions are "predominantly emotional"; "driven by feeling of anger"; proportional with that feeling; and generally more consistent with retributivism than consequentalism—what normative conclusions follow from such facts? As a practical matter, as Paul Robinson has argued, any proposed legal reform would be wise to take account of these facts.[41] Moreover, punishment decisions that deviated too far from the judgments of what most citizens think are fair may face a legitimacy problem. But these facts by themselves do not yet pose a normative challenge to retributivism per se.

The crucial move in presenting such a challenge is to link retributivism with deontology. "Deontologists," Greene contends, "argue that the primary justification for punishment is retribution"[42] and that "people's deontological and retributive punitive judgments are primarily

---

[39] Id. at 50.

[40] Id. at 51 ("Several studies speak to this question, and the results are consistent.")

[41] See Robinson, supra note 32.

[42] Greene, Secret Joke, supra note 6, at 50.

emotional."[43] Deontological judgments are produced by the "emotional" psychological process rather than the "cognitive" process, and consequentialist judgments are produced by the cognitive process. The cognitive process is more likely to involve "genuine moral reasoning," as opposed to the "quick," "automatic," and "alarm-like" deontological judgments produced by emotional responses. The supposed normative implications of this empirical information are to undermine deontology as "a rationally coherent moral theory";[44] an "attempt to reach moral conclusions on the basis of moral reasoning"; "a school of normative moral thought"; and as reflecting any "deep, rationally discoverable moral truths."[45] Rather, deontology is portrayed as merely an attempt to rationalize our emotional responses, which are based on, and may have developed evolutionarily because of, non-moral factors. The same goes for retributivism: "when we feel the pull of retributivist theories of punishment, we are merely gravitating toward our evolved emotional inclinations and not toward some independent moral truth."[46]

This purported neuroscientific challenge to retributivism is based on two conceptual mistakes. The first mistake is to equate retributivism with deontology. To the extent Greene assumes that retributivists about punishment are or must be deontologists about morality—or that retributivism as a theory of punishment necessarily depends on deontology—he is just plain

---

[43] Id. at 55.

[44] Id. at 72.

[45] Id. at 70-72.

[46] Id. at 72. Greene provides no arguments that utilitarianism or consequentialist judgments constitute or are the product of "a rationally coherent moral theory" or discover "deep, rationally discoverable moral truths." He does assert that "the only way to reach a distinctively consequentialist judgment . . . is to actually go through the consequentialist, cost-benefit reasoning using one's 'cognitive' faculties, the ones based in the dorsolateral prefrontal cortex." Id at 65. But the fact that one engages in explicit cost-benefit reasoning does not establish that the reasoning is the product of a coherent moral theory, much less that it discovers "deep" (or even shallow) moral truths. These further conclusions require the very types of philosophical arguments Greene decries when they are offered to support deontology or retributivism.

wrong. One may be a retributivist about punishment without being a deontologist about morality, and one may be a deontologist about morality without being a retributivist about criminal punishment.[47] The second mistake is to assume that retributivism entails the view that retributivist principles provide necessary and sufficient conditions for punishment. There are many coherent forms of retributivism that reject this assumption. For example, a retributivist theory may assert that the core retributivist idea of desert (1) provides a defeasible condition for punishment but concede that desert-based principles may be overridden by consequentialist considerations; (2) provides a necessary but not sufficient condition for punishment that would constrain consequentialist punishment decisions; or (3) justifies punishment but that consequentialist principles do so as well. Because of these two conceptual mistakes, Greene's argument does not go through. The neuroscientific facts to which he points do not undermine retributivism in all its forms.

Even though Greene's argument does not extend to all forms of retributivism, perhaps, he might reply, it does provide a plausible challenge to a limited subset of retributivist views. Specifically, his argument may challenge retributivist theories that meet two conditions: (1) the theory depends on a foundation of deontological morality, and (2) the decisions implied by this theory are correlated with neural activity in more "emotional" areas in the brain. But his argument does not effectively undermine even this subset of retributivist views. The argument would succeed only if there were reason to think that punishment decisions implied by this theory were somehow incorrect or unreliable. But this would presuppose some criteria by which we could establish whether particular decisions were correct or whether types of decision were

[47] See the discussion in Part I; Alexander & Moore, supra note 29.

reliable.[48] Greene provides no such criteria. He attacks deontology (and, by a loose extension, retributivism) for not having access to some "independent [moral] truth," but this is precisely the kind of access he would need to impugn the decisions implied by a retributivist theory. Nor is there any reason to think decisions implied by consequentialist theories of punishment would have better access to an independent moral truth.[49]

In sum, retributivism does not depend on a particular moral theory, much less on particular brain activity. The success or failure of retributivism does not depend on the success or failure of moral theories, and it does not depend on the areas of the brain associated with punishment decisions. Brain activity does not provide criteria for whether punishment decisions are correct or just, nor does the fact that retributivist decisions are associated with emotional decision-making provide evidence that the decisions are incorrect or unjust.

III. Neuroscience and Intuitions about Punishment

The second argument for neuroscience's purported challenge to retributivism is more sweeping and radical. Rather than focusing on the neural activity of those engaged in punishment decisions, this challenge focuses on the neural activity of criminals while committing criminal acts, and, indeed, the neural activity underlying all human action. Through such focus, Greene and Jonathan Cohen argue, "[n]euroscience will challenge and ultimately reshape our intuitive sense(s) of justice" and with it retributivism.[50]

---

[48] Moreover, even if there were some independent criteria by which to measure whether decisions are correct or reliable, it might turn out that engaging in cost-benefit analysis or consequentialist reasoning would lead to more mistakes. These would be open empirical questions that would depend on first having established normative criteria.

[49] See supra note 46.

[50] Greene & Cohen, supra note 6, at 208.

Before turning to the details of the argument, it is important to outline the familiar[51] philosophical positions regarding free will (or freedom of action) and physical determinism, on which Greene and Cohen rely. "Determinism" is the position that the world in its current state is "completely determined by (1) the laws of physics and (2) past states of the world," and that future states will be likewise so determined.[52] "Free will," as they define it, "requires the ability to do otherwise."[53] "Compatibilism" is the position that determinism, if true, is compatible with human free will.[54] "Incompatibilism" is the position that determinism and free will are incompatible and thus both cannot be true. Within incompatibilism, "hard determinism"

---

[51] Although familiar, these notions are anything but clear.

[52] Id. at 210 ("Given a set of prior conditions in the universe and a set of physical laws that completely govern the way the universe evolves, there is only one way that things can actually proceed.") Greene and Cohen acknowledge the existence of a certain amount of indeterminacy or randomness in the universe based on quantum effects, but they point out that this amendment adds little to the debate about how free will can emerge within the physical universe. Id. at 211. See also Peter van Inwagen, How to Think about the Problem of Free Will, 12 Ethics 327, 330 (2008) ("Determinism is the thesis that the past and the laws of nature together determine, at every moment, a unique future."); David Lewis, Are We Free to Break the Laws, 47 Theoria 112 (1981).

[53] Id. at 210. See also van Inwagen, supra note 52, at 329 ("The fee-will thesis is that we are sometimes in the following position with respect to a contemplated future act: we simultaneously have both the following abilities: the ability to perform the act and the ability to refrain from performing the act (This entails that we *have been* in the following position: for something we did do, we were at some point prior to our doing it able to refrain from doing it, able not to do it.)") Part of the confusion in discussions of free will stems from the fact that in arguments about whether an agent "can" or "has the power" to do something or to refrain from doing something, the terms "can" or "power" are ambiguous. As Anthony Kenny explains, in this context they may mean one of four different things: (1) natural powers (such as the ability of water to freeze) in which physical conditions may be sufficient for their instantiation; (2) abilities that depend for their exercise on an agent's wanting to exercise them; (3) opportunities to exercise one's abilities (one cannot swim if there is no water around); and (4) the presence of both an ability and an opportunity to exercise it. See Anthony Kenny, Freewill and Responsibility 30 (1978). The fourth sense is typically the relevant one when discussions of free will depend on whether an agent had the power to do otherwise. Id.

[54] Greene & Cohen, supra note 6, at 211. See also van Inwagen, supra note 52, at 330 ("Compatibilism is the thesis that determinism and the free-will thesis could both be true.") Note that the compatibilist need not take a stand on the empirical question of whether determinism is actually true. Rather, assuming the truth of determinism, the compatibilist is committed to the possibility that some human actions will be consistent with free will.

recognizes the incompatibility and denies free will; by contrast, "libertarianism" recognizes the incompatibility but denies determinism and accepts free will.[55]

Turning now to the details of the argument by Greene and Cohen, the first step in their challenge is to link the legitimacy of law with whether it "adequately reflect[s] the moral intuitions and commitments of society."[56] They note that while "current legal doctrine" (including criminal law and sentencing) may be "officially compatibilist," the intuitions on which such doctrine is based are "incompatibilist" and "libertarian."[57] Indeed, they contend that within "modern criminal law" there has been a "long, tense marriage" between "compatibilist legal principles" and "libertarian moral intuitions."[58] Neuroscience will "probably render the marriage unworkable" by undermining the moral intuitions: "if neuroscience can change those intuitions, then neuroscience can change the law."[59]

---

[55] Greene & Cohen, supra note 6, at 211-12.

[56] Id. at 213.

[57] Id. at 208. Given their preference for empirical data over philosophical arguments, it is curious how little empirical support Greene and Cohen provide for this claim that criminal-law doctrine is based on libertarian intuitions. They rely on two sources that, for different reasons, raise the possibility that brain damage (one source) or brain development in juveniles (the other source) may be relevant to criminal responsibility. See id at 213-17. As an empirical matter, however, non-legal actors appear to be "compatibilist" in their moral intuitions about particular cases. See Eddy Nahmias, Stephen Morris, Thomas Nadelhoffer & Jason Turner, Surveying Freedom: Folk Intuitions and Free Will and Moral Responsibility, 18 Philosophical Psychology 561 (2005). And legal doctrine in this area does not appear to depend on any tacit libertarian assumptions. See Stephen J. Morse, The Non-Problem of Free Will in Forensic Psychiatry and Psychology, 25 Behavioral Science & the Law 203 (2007); Peter Westen, Getting the Fly Out of the Bottle: The False Problem of Free Will and Determinism, 8 Buffalo Criminal L. Rev. 599 (2005). If Greene and Cohen are intent on debunking libertarian presuppositions in law perhaps a better target would be the doctrine in criminal procedure regarding the voluntariness of confessions and *Miranda* warnings, not criminal responsibility writ large. See Ronald J. Allen, *Miranda*'s Hallow Core, 100 Northwestern U. L. Rev. 71 (2006).

[58] Greene & Cohen, supra note 6, at 215.

[59] Id. at 215, 213.

The tension between legal doctrine and underlying moral intuitions is particularly acute with criminal punishment based upon retributivist principles. Retributivism and the doctrine it supports depend on notions of moral responsibility and "the intuitive idea that we legitimately punish to give people what they deserve based on their past actions."[60] Both "retributivism" and "moral responsibility," they contend, are incompatibilist, libertarian notions that rely on some kind of "magical mental causation"[61] within a "folk psychological system" of explaining human action. More specifically, retributivism depends on a notion of moral blameworthiness, and moral blameworthiness depends on a conception of folk psychology in which human actions are free from a deterministic physical universe. As Greene and Cohen put it, "folk psychology is the gateway to moral evaluation"[62] and [s]eeing something as an uncaused causer is a necessary but not sufficient condition for seeing something as a moral agent."[63]

The problem, as they see it, is that "hard determinism is mostly correct," folk psychology is "an illusion," and we are not "uncaused causers."[64] The notions of free will, moral responsibility, blameworthiness, and retributivist principles of punishment that depend on folk psychology and uncaused causation are therefore without a legitimate foundation. Neuroscience will help us to see the light by undermining "people's common sense, libertarian conception of free will and the retributivist thinking that depends on it, both of which have been shielded by the

---

[60] Id. at 210.

[61] Id. at 217.

[62] Id. at 220. They add: " To see something as morally blameworthy or praiseworthy . . . one has to first see it as 'someone,' that is, as having a mind." Id.

[63] Id. at 221.

[64] Id. at 221, 209.

inaccessibility of sophisticated thinking about the mind and its neural basis."[65] Once the folk-psychological illusion has been revealed, we can "structure our society accordingly by rejecting retributivist legal principles that derive their intuitive force from this illusion."[66]

How exactly will neuroscience do this? It will do so, they predict, by revealing the "mechanical nature of human action," along with the "when," "where," and "how," of the "mechanical processes that cause behavior." [67] As they acknowledge, this is not a new conclusion: "[s]cientifically minded philosophers have been saying this ad nauseam."[68] But the neuroscience will reveal this mechanical nature in a way that "bypasses complicated [philosophical] arguments," for it is one thing to hold your ground in the face of a "general, philosophical argument" but "quite another to hold your ground when your opponent can make detailed predictions about how these mechanical processes work, complete with images of the brain structures involved and equations that describe their functions."[69]

To illustrate how this might work, they present the following hypothetical. Imagine a group of scientists who create an individual ("Mr. Puppet") who engages in criminal activity. At Mr. Puppet's trial, the lead scientist explains his relationship to Mr. Puppet as follows:

> I designed him. I carefully selected every gene in his body and carefully scripted every significant event is his life so that he would become precisely what he is today. I selected

[65] Id. at 208.

[66] Id. at 209.

[67] Id. at 217.

[68] Id. at 214.

[69] Id. at 217.

> his mother knowing that she would let him cry for hours and hours before picking him up. I carefully selected each of his relatives, teachers, friends, enemies, etc. and told them exactly what to say to him and how to treat him.[70]

According to Greene and Cohen, Mr. Puppet is guilty according to the law if he was rational at the time of his actions, which, they assume, he was. However, they conclude that given the circumstances of his creation "intuitively, this is not fair."[71] It is not fair, they contend, because "his beliefs and desires were rigged by external forces, and that is why, intuitively he deserves our pity more than our moral condemnation."[72] What neuroscience will reveal—without the need for recourse to philosophical argument—is that all criminal defendants (and indeed all humans) are like Mr. Puppet in the relevant respects. Although not designed by scientists, our beliefs, desires, and "rational" actions are all the rigged product of external forces beyond our control (some combination of genes, history, culture, and perhaps randomness). If Mr. Puppet is not morally responsible, then no one else is either.

Neuroscience will reveal the mechanical nature of our actions with examples like the following:

> Imagine, for example, watching a film of your brain choosing between soup and salad. The analysis software highlights the neurons pushing for soup in red and the neurons pushing for salad in blue. You zoom in and slow down the film, allowing yourself to trace the cause-and effect relationships between individual neurons—the mind's

[70] Id. at 216.

[71] Id. at 216.

[72] Id. at 216.

> clockwork revealed in arbitrary detail. You find the tipping-point moment at which the blue neurons in your prefrontal cortex out-fire the red neurons, seizing control of your pre-motor cortex and causing you to say, "I will have the salad, please."[73]

What goes for the soup-or-salad choice, also goes for the choice of whether to murder, rape, assault, or steal.

Greene and Cohen do not see these neuro-revelations as the end of criminal punishment, however. They note that the "law will continue to punish misdeeds, as it must for practical reasons,"[74] and that "if we are lucky" our retributivist reasons for punishment will give way to consequentialists reasons,[75] because "consequentialist approaches to punishment remain viable in the absence of common-sense free will."[76] Under a consequentialist punishment regime, legal doctrine may, for deterrence purposes, make many of the same distinctions it does today (for example, regarding infancy and insanity), "but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless."[77] They end with rhetorical flourish: "the law deals firmly but mercifully with individuals whose behavior is obviously beyond their control. Some day, the law may treat all convicted criminals this way. That is, humanely."[78]

---

[73] Id. at 218.

[74] Id. at 218.

[75] Id. at 224.

[76] Id. at 209.

[77] Id. at 218.

[78] Id. at 224.

There are a number of problems with this argument. Carefully examining each of these distinct problems will reveal how little the neuroscience of human action *tout court* bears on the normative project of justifying criminal punishment on the basis of moral blameworthiness or desert. Each problem by itself is sufficient to raise doubts about the conclusions Greene and Cohen draw regarding retributivism; collectively they illustrate why their conclusions ought to be rejected.

The first problem with the argument is the assumption that the intuitions of most people necessarily answer the normative questions of whether criminal punishment is justified or how it ought to be distributed. Although lay intuitions may be relevant to reform, and some agreement between punishment and lay intuitions may be *necessary* for the legitimacy of such punishment, accord with the intuitions of most people is not *sufficient* to justify punishment decisions. It is possible for widely shared intuitions about what is just punishment to be mistaken. Thus, even if neuroscience were to cause a significant shift away retributive intuitions (as they predict),[79] it begs the question to assume this shift would lead to more just (or more unjust) punishment decisions. The key issue is whether neuroscience contributes evidence that provides epistemic support for arguments concerning compatibilism versus incompatibilism, moral blameworthiness, and just punishment.

The second problem with their argument is that the neuroscientific evidence does not provide this epistemic support. As Greene and Cohen appear to concede with their off-hand

[79] For preliminary empirical support for the idea that exposure to deterministic ideas may reduce punishment for retributivist reasons in certain circumstances, see Shariff, Greene & Schooler, supra note 6. In other circumstances, however, abandoning retributivism because of deterministic thinking may cause more, not less, punishment. For example, consider current practices of indefinite "civil commitment" of convicted sex offenders after they have completed their prison sentences. The assumption that getting rid of retributivism will reduce punishment neglects the constraining or limiting effects retributivist thinking may provide.

references to and dismissal of "complicated [philosophical] arguments,"[80] neuroscience adds nothing new to extant conceptual arguments for or against compatibilism, incompatibilism, or hard determinism.[81] If this is so, and the presence of neuroscientific information causes people to form and hold new beliefs about these positions, then the neuroscience is persuading people for psychological reasons other than the epistemic support it provides. As in other contexts, the presence of neuroscientific information may be causing people systematically to draw faulty or unsupported inferences rather than true or justified ones.[82] In other words, the effects that Greene and Cohen predict may be widespread cognitive mistakes. The neuroscientific information may cause people to draw problematic (or philosophically dubious) inferences regarding the issues of free will, determinism, and responsibility. Perhaps Greene and Cohen would respond that this causal effect is at least pushing people toward the correct positions, albeit for the wrong reasons. But this presupposes that their assumption that retributivism depends necessarily on libertarianism is correct (for reasons unrelated to neuroscience). This takes us to a third problem with their argument.

The third problem is that their assumption is mistaken. It is not the case that retributivism depends necessarily on a metaphysically problematic version of libertarian imcompatibilism. Greene and Cohen assume that retributivism—and indeed all moral blame and

---

[80] See Greene & Cohen, supra note 6, at 217.

[81] Indeed, both the "Mr. Puppet" and "soup/ salad" examples are consistent with a variety of different positions on these issues.

[82] The causal role played by neuroscientific evidence in drawing inferential conclusions need not necessarily be a justifactory role. See Jessica R. Gurley & David K. Marcus, The Effects of Neuroimaging and Brian Injury on Insanity Defenses, 26 Behavioral Sciences & the Law 85 (2008); David P. McCabe & Alan D. Castel, Seeing is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning, 107 Cognition 343 (2008); Deena Skolnick Weisberg et al., The Seductive Allure of Neuroscience Explanations, 20 J. Cognitive Neuroscience 470 (2008). Whether the inferences are justified or not will depend on other criteria (including philosophical and conceptual arguments) beyond what caused them.

praise—must be built on a foundation of actions by "uncaused causers." But a retributivist can coherently reject the notion of uncaused causers and still allow for moral judgments. Even in a world of physical determinism, moral desert may be grounded in the control people have over their actions through the exercise of their practical rationality.[83] If people act for reasons—more generally, if they act on the basis of their beliefs, desires, and other mental states—then we can blame or praise their actions (in light of their mental states),[84] so long as they had the ability and the opportunity to act differently.[85] Indeed, in their appeal to consequentialist justifications for punishment based on deterrence, Greene and Cohen appear to concede this type of responsiveness to reason: deterrence works precisely by affecting the practical rationality of potential offenders, by giving them a reason to refrain from criminal activity that (ideally) outweighs their reasons for criminal activity. Sufficient control over one's actions in light of

---

[83] Rational control does not imply "uncaused causation." It implies that people have the ability and the opportunity to act in accord with their mental states. Determinism is consistent with the idea that wants, beliefs, desires, intentions, and other mental states affect behavior. For a discussion that develops these points, see Anthony Kenny, Freewill and Responsibility 32 (1978) ("whatever story the physiological determinist tells about my present physiological state it must contain a proviso that my brain state would be different from what it now is if I wanted something different from what I now want.")

[84] This does not assume that actors are always conscious of their mental states or that the mental states necessarily precede actions. In some case there may be no unique mental state that may be distinguished from the action that manifests the mental state (e.g., a want, knowledge, or an intention).

[85] Determinism is consistent with the idea that people possess the ability and the opportunity to do otherwise. See Kenny, supra note 83, at 34 ("it does not follow from determinism that agents always lack the opportunity and ability to do otherwise than they do. Consequently it does not follow from determinism that it is unfair to hold people responsible for their actions.") How can ability and opportunity be consistent with determinism? Consider first ability. Possessing an ability (e.g., to ride a bicycle) depends on whether one satisfies the criteria for possessing the ability. These criteria include successfully exercising this ability when one wants to (and has the opportunity to) and refraining when one does not want to (and the opportunity to refrain). Such criteria can be fulfilled even if one does not exercise the ability on a particular occasion. Second, consider opportunity. One has the opportunity to act (or not to act) if conditions external to the person are not forcing or preventing the exercise of the ability on a particular occasion. But why aren't one's brain states forcing the person to act one way or another and thus depriving them of the opportunity to do otherwise? Again, we can assume that if one's wants had been different (e.g., to ride a bicycle or not), then one's brain states also would have been different. See supra note 83. It would be a quite different story if one's brain states caused one to ride a bicycle (or not) when one wanted to do the contrary. But in such circumstances we would have a breakdown of the type of rational control on which responsibility depends.

one's practical rationality is sufficient to ground moral desert, regardless of whether the same actions may be explained in purely physical (i.e. non-mental) terms. In other words, one can coherently be a compatibilist and a retributivist, a combination that is consistent with current law. To suppose otherwise would be a mistake, regardless of how many people think so, and regardless of what neuroscience shows.[86]

Their example of Mr. Puppet will help to illustrate this conclusion. Suppose Mr. Puppet has robbed a bank. Let's assume determinism is true and that we must decide whether to hold Mr. Puppet responsible for his actions. Assume further than Mr. Puppet is responsible only if he acted freely in robbing the bank, in the sense that he had the ability and the opportunity to not rob the bank. We ask him why he did so and he says, "I wanted the money." We might say the money (or his wanting the money) caused him to rob the bank, but this would not negate moral blame.[87] Presumably, Mr. Puppet had the ability to refrain from robbing the bank, in the sense that his mental states (his beliefs and desires) played some causal role in his conduct and he was responsive to reasons for and against his conduct at the time of his actions.[88] His ability to act or not in robbing the bank was thus distinct from someone sleepwalking or insane at the time. If,

---

[86] Rather than causing people to abandoning retributivism and libertarian free will, perhaps increased neuroscientific knowledge will instead cause people to abandon confused views about free will and its relationship with responsibility? An important role for philosophy in this endeavor is helping to integrate this increased knowledge coherently into the conceptual schemes we use to explain human behavior and the world.

[87] Causation, even abnormal causation, does not necessarily equal excuse. See Morse, supra note 57. Moreover, the wanting need not be a distinct event that precedes robbing the bank; it may be manifested in the robbing itself.

[88] This is based on the assumptions by Greene and Cohen that Mr. Puppet: "is as rational as other criminals and, yes, it was his desires and beliefs that produced his actions." Id. at 216. They also raise the possibility of defining "rationality" in neurocognitive rather than behavioral terms. Id. at 224 n. 3. But this would either be changing the subject (i.e., we would no longer be talking about what we currently mean by rationality) or incoherent as an explanation of rationality as currently conceived. People, not brains, behave rationally (or not). It is an instance of the "mereological fallacy" (i.e., mistakenly ascribing attributes to parts that make sense only when ascribed to the whole) to assume rationality may refer to states of the brain. See M.R. Bennett & P.M.S. Hacker, Philosophical Foundations of Neuroscience (2003).

for example, Mr. Puppet learned the police were waiting inside the bank, we can presume that he would respond to this information and (given his beliefs and desires to have the money and not go to jail) abandon his plan to rob the bank—thus exercising this ability. By contrast, a sleepwalker or an insane person may not have the ability to respond to this information in a similar manner. Possessing an ability does not require exercising it whenever possible, so even though Mr. Puppet did not exercise it in the deterministic world in which he robs the bank, this does not mean he lacked the ability to do otherwise.[89]

But did Mr. Puppet have an opportunity to do otherwise? In an important sense—yes. No external forces were coercing Mr. Puppet when he acted.[90] We can also assume that if Mr. Puppet did not want the money, his brain states would be different from his brain states when he wanted the money and robbed the bank and, thus, he would have acted differently.[91] Therefore, whatever Mr. Puppet's neurological and other physical states are in the deterministic world, it is not case that if Mr. Puppet did not want to rob the bank, these physical states would cause him to

---

[89] See supra notes 83 and 85. Mr. Puppet had the ability to refrain from robbing the bank if he could exercise that ability when he wanted to (and when there is an opportunity to do so).

[90] Although, under some formulations of the hypothetical, we might have grounds for inferring that the scientists who designed Mr. Puppet coerced his behavior. Some type of coercion by third parties is typically what people mean by the claim that one's action was not free. See Nahmias et al., supra note 57; van Inwagen, supra note 52, at 329 ("['Free will's'] non-philosophical uses are pretty much confined to the phrase 'of his/her own free will' which means 'uncoerced.'")

[91] Similarly, if one wanted soup rather than salad in the previous example, we can assume one's neurons would have been different. To suppose otherwise, Greene and Cohen would have to defend much stronger claims than they do: namely, that (1) there is a one-to-one correspondence between brain states and particular mental states, and (2) the relationships between various mental states and between mental states and actions are governed by the same physical laws that govern brain states (or are reducible to those laws). They do not defend either claims, cf. Greene & Cohen, supra note 6, at 225 ("we do not wish to imply that neuroscience will inevitably put us in a position to predict any given action based on a neurological examination"), and neither claim necessarily follows from determinism. Plausible positions that reject either claim are consistent with physical determinism. See Donald Davidson, Mental Events, in Essays on Actions and Events 207 (2001); Richard Rorty, The Brain as Hardware, Culture as Software, 47 Inquiry 219, 231 (2004). Greene and Cohen do, however, assert the stronger claim that folk psychology is an illusion. See Greene & Cohen, supra note 6, at 209. But this denies rather explains the causal role of mental states. We turn to this aspect of the argument below.

do so anyway or deprive him of the opportunity to adhere to the law. Once again, compare Mr. Puppet with a person who cannot exercise this control. Suppose a person cannot bring her actions to conform to her desires, goals, plans, and intentions, or, for a variety of reasons, cannot control her bodily movements.[92] It is precisely in such cases that we withhold judgments of moral blame—and indeed often do not even consider such movements to be "actions" at all—and the criminal law withholds punishment.

The consistency between moral judgment and determinism becomes even clearer when focusing on acts of moral praise. Suppose that, instead of a criminal act, Mr. Puppet commits an act of heroic bravery or kindness—for example, sacrificing himself in some way to save a stranger. Does his heroic or kind act cease to be morally praiseworthy if it takes place in a deterministic physical world and he is the product of his genes and upbringing? We think not. As with moral blame, what matters is whether Mr. Puppet can act for reasons and can exercise control over his actions on the basis of those reasons. Did he have the ability and opportunity to do otherwise and act voluntarily in performing this praiseworthy act? Contrast this with the bodily movements of someone that were not within that person's rational control. For example, if someone in a state of epileptic seizure or while fainting engages in bodily movements that turn out to somehow save a third party, has the person acted heroically or bravely? Do they deserve moral praise? Have they acted at all? As with moral blame, we think the distinction here is plain as well. (We also doubt that most people would be persuaded by neuroscience to think

[92] Examples of the latter might include some cases of "alien hand" syndrome or "utilization behavior." See Iftah Biran et al., The Alien Hand Syndrome: What Makes the Alien Hand Alien? 23 Cognitive Neuropsychology 563 (2006).

otherwise, but this is an empirical question for "experimental philosophy."[93]) Mr. Puppet deserves praise for his morally good acts, along with any other praiseworthy accomplishments, and blame for morally bad acts, when he had the ability and opportunity to do otherwise.

To suppose moral praise or blame require an uncaused causer is to miss (or misconstrue) the normativity in human action. Our normative judgments about human actions are not inconsistent with physical determinism; they require, at a minimum, that our bodily movements be *human actions* and not mere bodily movements—that is, that they are explainable based on the actor's mental states[94]—and that these actions meet or fail to meet various moral standards or criteria, not that we be "uncaused causers." Greene and Cohen deny normativity on this level, however, along with the distinctions we have been drawing with Mr. Puppet, by arguing that they are based on the "illusion" of folk psychology. The "illusion," in their view, is the idea that our (and Mr. Puppet's) mental states (beliefs, desires, and intentions) are real and play a causal role in behavior. They think that either there are no such things or, if there are such things, they do no causal work (are "epiphenomenal"). This leads to a fourth problem with their argument.

The fourth problem is that neither determinism in general nor neuroscience in particular undermines folk psychology in the ways they suppose. It is important to be clear about the nature of their claim here and the role that neuroscience purports to play in it. Their argument is based on the following premises and inferences: (1) retributivist punishment based on moral desert requires, at a minimum, that people punished had some control over their actions; (2) to

---

[93] Similar to the "Knobe effect" with regard to ascriptions of intentions, see Joshua Knobe, Intentional Actions and Side Effects in Ordinary Language, 63 Analysis 190 (2003), subjects might distinguish between good acts and bad acts for reasons other than the relationship between determinism and free will.

[94] See G.E.M. Anscombe, Intention (1957).

have such control they must have been able to act or refrain from acting; (3) the ability to act or refrain from acting requires that their mental states played a causal role in regulating their behavior; (4) but these mental states do not exist or do no causal work; (5) thus they had no such control over their actions; (6) thus retributive punishment is unjustified.[95] Neuroscience, it is claimed, will support premise (4) by illustrating that behavior is determined by one's neurological states. But this does not follow. Few who claim that mental states exist and play a causal role would deny that mental states have underlying neurological correlates. Thus, the simple fact that mental states have accompanying brain states does not render the former illusory or epiphenomenal. Moreover, as an empirical matter, some mental states (e.g., intentions) do appear to play a causal role in a manner that would be impossible if folk psychology were an "illusion."[96] This alone should be enough to neutralize this radical claim from Greene and Cohen, but there are deeper and more problematic issues with their claim.

Reflect for a moment on what it would mean for folk psychology to be an illusion (i.e., for it to be false that there are no such things as beliefs, desires, wants, fears, knowledge, intentions, plans). One obvious implication is that there is no difference between us (as well as Mr. Puppet) and someone engaged in bodily movements caused by a seizure or conditions beyond their "rational" control. A second implication is that psychological explanations would

---

[95] Note that this aspect of the argument does not actually depend on an implausible notion "uncaused causers." A stronger argument could be constructed involving this notion, but we construe the argument in its weaker form to make it as plausible as possible for purposes of our discussion.

[96] For an overview of some of the relevant literature, see Peter M. Gollwitzer & Paschal Sheeran, Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes, 38 Advances in Experimental Social Psychology 69 (2006). For a discussion of the relevance of these studies to debates about free will, see Alfred R. Mele, Effective Intentions: the Power of Conscious Will (2009). See also Mario Beauregard, Mind Really Does Matter: Evidence from Neuroimaging Studies of Emotional Self-Regulation, Psychotherapy, and Placebo Effect, 81 Progress in Neurobiology 218 (2007).

be false, and there would be nothing real (or nothing that affects behavior) for psychology to explain. A third implication is that human action would indeed cease to be "human" or "action" as we currently conceive of these notions.

Now, notice how self-defeating this is as a challenge to *retributivism* and as a defense of *consequentialist* criminal punishment. Recall, they argue that we will still have to punish for practical reasons, for example, to deter future crime. Well, why? Do we *want* to deter crime? Do we *believe* or *know* punishment will deter crime? Will we therefore *choose* certain forms of punishment over others?[97] Notice also that consequentialist justifications for punishment also involve folk psychological concepts; deterrence works by affecting the practical reasoning of potential criminals. Do potential criminals *believe* punishment will follow if they commit crime; do they not *want* to be punished; and therefore will they *choose* to not commit crimes? Their arguments presuppose the existence of the very entities they purport to deny to support their conclusions.

But, they suggest, perhaps we can have it both ways. Perhaps we can invoke these entities and concepts as a general matter for practical reasons (for example, deciding who has committed a criminal act), but that for the special case of determining criminal punishment we can fall back on the conclusion that retributive punishment is unjustified because no one is *really* responsible (because they didn't really *act* at all.) Well, why would we do so? Why should we *want* to do so or *believe* it is a good *idea* to do so? If folk psychology is an illegitimate basis on which to ground and justify criminal punishment, it is not clear why Greene and Cohen think it

[97] See, e.g., Shariff, Greene & Schooler, supra note 6, at p.__: "Giving up on free will could make people more willing to commit moral transgressions and less willing to seek punishment for transgressors as an end in itself." In the absence of folk psychology, we have no idea what "willing" even refers to in this sentence?

would be an appropriate basis for singling out people to punish in the first place (even for "practical reasons"). If beliefs play no causal role in behavior, not only can they not be used as a basis to justify criminal punishment on retributivist grounds; they also cannot be used to single out criminal from non-criminal behavior, to create justifications or excuses, or to justify punishment on consequentialist grounds.[98]

Finally, there is a fifth problem with their argument. Even if they are right in their predictions; even if people are persuaded based on neuroscience to believe in hard determinism; even if they come to *believe* that folk psychology is an illusion and they don't have beliefs; and even if they therefore abandon retributivism as a basis for justifying punishment; Greene and Cohen are wrong to suppose that we would be "lucky" and punishment would necessarily be more "humane." Although some recent and interesting experimental work by Greene and colleagues suggests that subjects recommended less punishment when abandoning retributivist reasons for punishment,[99] a brief history of actual criminal-sentencing practices in the United States over the last thirty years suggests the reverse to be true. Abandoning retributivist rationales for punishment in favor of deterrence, incapacitation, and general crime control has given us three-strikes laws, harsh sentences for drug crimes, prosecuting juveniles as adults, strict liability crimes, proposals to abolish the insanity defense, and the felony-murder rule.[100] It has also given us proposals for the indefinite "civil commitment" of criminals. Indeed, more widespread belief that criminals cannot stop and are "determined" to continue their criminal

---

[98] See Kenny, supra note 52, at 93 ("A legal system which took no account of states of mind would be as chimeric as it would be abhorrent.")

[99] See Shariff, Greene & Schooler, supra note 6.

[100] See Paul H. Robinson, Owen D. Jones & Robert Kurzban, Realism, Punishment, and Reform, 77 U. Chi. L. Rev. 1611, 1630 (2010).

behavior does not appear to be a recipe for more compassionate and humane punishment. Moreover, psychologically persuasive but epistemically dubious neuroscience may only exacerbate rather than alleviate this problem.[101] We share what we believe to be the sentiment of Greene and Cohen that criminal punishment ought to be more humane, but we do not believe that the way to get there is by denying our shared humanity. Instead, we concur with Anthony Kenny that "[a] legal system which took no account of states of mind would be as chimeric as it would be abhorrent."[102]

We close with a final point. Imagine a group of open-minded policymakers faced with the task of constructing a justified system of legal punishment; they decide to listen to and take seriously the arguments of Greene of Cohen regarding retributivism, hard determinism, and neuroscience. At the end of the day, they could evaluate the various claims and the reasons for them; deliberate about the various avenues open to them and the benefits and costs of each; and then choose a course of action they think is justified or more justified than the alternatives. Or they could simply sit back and wait for their neurons to make the decision for them. For the normative project of justifying criminal punishment, the distinction matters a great deal to whether the criminal punishment that follows is justified and legitimate.[103] From the neuro-reductionist perspective of Greene and Cohen, however, this difference ultimately does not matter (just as it does not matter at the level of criminal behavior)—if the mental states, practical

---

[101] See supra note 79. This prediction was made recently by Richard Sherwin. See Richard K. Sherwin, Law's Screen Life: Criminal Predators and What to Do About Them, in Imaging Legality: Where Law Meets Popular Culture (Austin Sarat, ed., forthcoming).

[102] Kenny, supra note 83, at 93.

[103] And it appears to matter in a self-defeating way for Greene and Cohen, who apparently think punishment should proceed for "good" consequentialist reasons and not "bad" retributivist reasons.

rationality, and choices of criminal defendants ultimately do not matter in whether their actions are blameworthy or praiseworthy, justified or unjustified, then the same goes for the lawmakers who decide how and when to distribute punishment. If it is just neuronal activity all the way down, regardless of the illusory mental states that seem to be involved, then it does not matter why anyone chooses to engage in criminal punishment or how they go about doing so.[104] The same goes for theorists engaged in the normative project of critiquing and defending possible policies regarding the distribution of criminal punishment. If so, then one wonders why they bothered to make the argument.

[104] Notice also the deep irony between this argument by Greene and Cohen and the argument by Greene considered in Part II. There, recall, it mattered a great deal which brain processes were correlated with the decision to punish ("emotional" or "cognitive") as well as whether the punisher went through an explicit consequentialist/ utilitarian/ cost-benefit reasoning process. The irony: this distinction between types of mental states and the normative conclusions that follow from the distinction presuppose the existence and significance of the very types of mental activity (and folk-psychological explanations) the argument in this Part asserts is an "illusion."