Pace University
DigitalCommons@Pace

Pace Law Faculty Publications | School of Law

3-25-2010

Free Will Ideology: Experiments, Evolution and Virtue Ethics

John A. Humbach
Pace University School of Law, jhumbach@law.pace.edu

Follow this and additional works at: http://digitalcommons.pace.edu/lawfaculty

Part of the Criminal Law Commons, and the Law and Society Commons

Recommended Citation
Humbach, John A., "Free Will Ideology: Experiments, Evolution and Virtue Ethics" (2010). *Pace Law Faculty Publications.* Paper 636.
http://digitalcommons.pace.edu/lawfaculty/636

This Article is brought to you for free and open access by the School of Law at DigitalCommons@Pace. It has been accepted for inclusion in Pace Law Faculty Publications by an authorized administrator of DigitalCommons@Pace. For more information, please contact cpittson@law.pace.edu.

Free Will Ideology:

Experiments, Evolution and Virtue Ethics

by John A. Humbach, Pace Law School
White Plains, New York
jhumbach@law.pace.edu
citation: http://ssrn.com/abstract=1428002

Table of Contents

I. 'Deserved' and Undeserved.
II. Justifying Infliction.
III. Evidence For and Against the Existence of Free Will.
a. Our introspective knowledge of free will.
b. Three experiments.
c. Discussion of the experimental evidence.
IV. Free Will and Evolution.
V. Free Will, Virtue-Ethics and 'Compatibilism.'
a. Virtue ethics as a substitute for 'free will.'
b. Virtue ethics as the real 'compatibilism.'
VI. A Few Last Words on Free Agency.
VII. Conclusion.

Summary of Main Points:

1. Hurting people is presumptively wrong, therefore…

2. It falls on those who would inflict human suffering to supply a moral justification for the hurt they cause.

3. Before "free will" can be invoked to justify inflictions of human suffering, its existence must be, at the very least, sufficiently well established to count as a reasonable basis for action (borrowing the usual legal standard).

4. The current state of the experimental evidence tends, if anything, to falsify the free will hypothesis and, at any rate, falls far short of establishing it to any reasonable degree of certainty.

5. In addition, as a matter of natural selection, free will could never have evolved under ordinary biological pressures.

6. In sum, the weakness of the case supporting the free will hypothesis prevents it from serving as a reasonable (*i.e.,* morally sufficient) basis for action.

7. Compatibilist theories shift the focus of "free will" to freedom of "action," but they do not solve or avoid the problem of justification. On the contrary, they seem on closer analysis to be merely a repackaging of an ancient form of virtue-ethics under which people are deemed to deserve to suffer because they are what they are.

Free Will Ideology:

Experiments, Evolution and Virtue Ethics

by John A. Humbach, Pace Law School
White Plains, New York

citation: http://ssrn.com/abstract=1428002

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

I.

'Deserved' and Undeserved. Most agree that human suffering is a serious matter, but many draw a sharp distinction between innocent suffering and inflictions that are 'deserved.' In turn, some concept of free will is normally invoked as the linchpin of just deserts. Before you can blame a person for being bad, you must first suppose there was a choice to be good.[2] It is when people *choose* to do wrong that their conduct is deemed culpable and deserving of punishment.[3] Indeed, in situations where free choice seems absent, such as duress or extreme necessity, both the legal tradition and morality tend to excuse or even justify actions that harm others.[4] For an act to be a crime, it has to be 'voluntary.'[5]

---

[1] THE BHAGAVAD GITA 3:27-28 (Juan Mascaró, trans. 1962) (*c.* 500 B.C.).

[2] "C'est ôter toute moralité à ses actions que d'ôter toute liberté à sa volonté." Jean-Jacques Rousseau, quoted in André Sénik, *Déterminisme et liberté: l'interminable débat?*, in LA RESPONSIBILITÉ 34, 35 (Monique Vacquin, ed., 1996), which also quotes René Descartes. *Cf.* STANFORD ENCYCLOPEDIA OF PHILOSOPHY, *Compatibilism* 1.1 ("For the most part, what philosophers working on this issue [of free will] have been hunting for, maybe not exclusively, but centrally, is a feature of agency that is necessary for persons to be morally responsible for their conduct."), available at http://plato.stanford.edu/entries/compatibilism/; Deborah W. Denno, *Crime and Consciousness: Science and Voluntary Acts*, 87 MINN. L. REV. 269, 275 (2002) ("Doctrinally, all criminal liability depends on one 'fundamental predicate': A defendant's guilt must be based on conduct and that conduct must include a 'voluntary act' or omission to engage in a voluntary act that the defendant was capable of performing").

[3] *See* René Descartes: "Il n'y a que ce qui depend de la volonté qu'on ait sujet de récompenser ou do punir." Oeuvres Philosophiques, quoted in Sénik, *supra* note 2, at 34. Nietzsche characterizes this notion as an 'extremely late' invention to justify the real reason for retribution, which he described as 'anger over some wrong which has been suffered.' See FRIEDRICH NIETZSCHE, ON THE GENEALOGY OF MORALITY 43 (Keith Ansell-Pearson ed., 1994).

A biological (evolutionary) explanation for the urge to retribution was outlined in my earlier paper *The Humane Principle and the Biology of Blame (Evolutionary Origins of the Imperative to Inflict),* in PROCEEDINGS OF 3RD ANNUAL GLOBAL CONFERENCE ON PERSPECTIVES ON EVIL AND HUMAN WICKEDNESS 145-62 (2003), available at http://www.law.pace.edu/jhumbach/TheBiologyofBlame.doc (the urge to retribution arose in the evolution as a psychological prompt to strike back against encroachments by others, and has become progressively dressed up in confabulations about a non-existent 'free' will.

[4] *See, e.g.,* United States v. Contento-Pachon, 723 F.2d 691 (9th Cir. 1984); State v. Toscano, 378 A.2d 755 (N.J. 1977); MODEL PENAL CODE §§ 2.09 & 3.02 (1962) duress and choice of evils. In less extreme situations, where free choice is clouded by 'provocation,' a partial excuse may be given. See, e.g., State v. Thornton, 730 S.W.2d 309 (Tenn. 1987) (finding wife in adulterous act). *See generally* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 289-321 (5th ed. 2009).

[5] *See, e.g.,* MODEL PENAL CODE §2.01 (1962); Denno, *supra* note 2.

At the core of this reasoning is the assumption that free will exists as a matter of *fact*. What if, however, free will does not in fact exist? What if people do not freely choose what they do but, rather, every human action is the mechanical result of causes and effects that reach back, in a continuous succession, indefinitely? What if, in other words, human actions are just like all of the other events that occur throughout the universe?[6]

Free will plays a crucial role in justifying punishment and other moralistic aggression. But if free will does not in fact exist, then a cherished rationale for inflicting human suffering (*viz.* the idea that some people 'deserve' to suffer) would fail. The problem is that the factual existence of free will is, to say the least, contestable.

The arguments for and against free will have a long history,[7] and the debate will probably be forever inconclusive.[8] Nonetheless, even while the factual existence of free will continues to be contested, the ideology of free will begets enormous human suffering. The mere belief that free will exists is a veritable engine of evil.

This essay maintains that the ideology of free will does harm, that it is a beguiling vehicle to bring suffering to the world.[9] The foundational claim is that hurting people is presumptively wrong; that is to say, inflictions of human suffering are immoral unless they are *justified*—with the grounds of justification being shown to a reasonable certainty. A justification that is based only on uncorroborated assertions or beliefs, no matter how sincere, is not adequate to overcome the presumption that hurting people is wrong.

We will begin, in Part II, with a consideration of what proof is needed to justify intentional inflictions of human suffering. Then we turn, in Part III, to a review of the weakness of the largely *introspective* evidence for the freewill ideology, followed by a brief discussion of some experiments that raise serious doubt about the fact of free will. Part IV presents an evolutionary case against the factual existence of free will, namely that, under the normal biological pressures of natural selection, a capacity for free will could never have evolved. Finally, in Part V, we will look at an alternate and less ambitious conception of free will (called 'compatibilism') that has been proposed, in various guises, to avoid difficulties posed by a more robust version of free will ideology. The argument will be made that, though compatibilist strategies may escape the case against genuine freedom of will,

---

[6] Certain kinds of events among quantum particles are probabilistic rather than strictly determinant, but being thus indeterminant does not make them "free." *See infra* note 40.

[7] *See* Sénik, *supra* note 2. Arthur Schopenhauer provided a useful summary of the debate up to his own time in his prize essay. ARTHUR SCHOPENHAUER, THE FREEDOM OF WILL.56-80 (1841: ERIC F. PAYNE, TRANS. 1999); BENJAMIN LIBET *et al.*, THE VOLITIONAL BRAIN: TOWARDS A NEUROSCIENCE OF FREE WILL 45-55 (1999); TED HONDERICH, HOW FREE ARE YOU? (1993).

[8] Roughly speaking, the indemonstrability problem seems to be partly due to the fact that free will is essentially a species of causality—an 'original' cause. And, as Hume pointed out, causality itself cannot be empirically or logically shown to exist, even in principle. DAVID HUME, AN ENQUIRY CONCERNING HUMAN UNDERSTANDING 58-95 (1748: 1988); *see also* DAVID HUME, A TREATISE OF HUMAN NATURE 69-94 & 173-76 (1737: L.A. Selby-Bigge, ed. 2d ed. 1978). If causation must be accepted à priori, *per* Kant, then free will cannot be empirically known except by starting with an à priori premise. But what premise?

[9] *Cf.* "We need to hold ourselves responsible for what we say, recognizing that our words, if believed, can have profound effects for good or ill." DANIEL C. DENNETT, FREEDOM EVOLVES 17 (2003) (arguing, however, that there exists a kind free will that is sufficient to justify punishment).

they still offers only a dubious 'character' rationalization—akin to value ethics—for infliction of human suffering.

II.

Justifying Infliction. The foundational claim from which this essay proceeds is that the infliction of human suffering is presumptively wrong. Now this is admittedly an arguable premise, and the possibility cannot be absolutely excluded that hurting people is either presumptively right or, at least, morally neutral.[10] For any who may actually take such a position, I can only say that this essay is not addressed to you. But for those who agree that hurting people is presumptively wrong, the case against the free will ideology runs as follows:

To inflict human suffering is morally repugnant unless there is an adequate justification. This is what it means to say that hurting people is presumptively wrong. Examples of adequate justification are potentially far ranging, including concerns as disparate as providing dental care, acting in self-defense, or preventing an even greater harm.[11] When harm is an inadvertent byproduct of ordinary legitimate activities, adequate justification might lie simply in the fact that the harm occurred despite reasonable care.[12] But in no case is it enough merely to *assert* that a justification exists. The presumption that hurting people is wrong means that the countervailing justification must be *shown.*

Moreover, if a putative justification depends on the existence of empirical facts (as most do), those facts cannot be merely conjectural or hypothetical. To show that the moral justification exists, the supporting empirical facts must also be demonstrated—for example, by showing that the painful dental procedure was medically reasonable, that the means of self-defense were necessary or that the other harm sought to be prevented was actually a greater one. If the facts underlying a purported

---

[10] Even if few today would admit believing that hurting people can be presumptively right, it is not so very long ago that some seriously contended that hurting or even annihilating large classes of people is not *necessarily* wrong. The situation in Europe around the early middle of the last century or in North America during phases of the Indian Wars or slavery might be cited as cases in point.

A particularly important modern possibility is the argument that hurting people cannot really be said to be wrong because right and wrong are purely relative and do not 'really' exist at all, . In this essay, however, I am sidestepping the question of whether right and wrong really exist or whether morality is indeed 'part of the fabric of the universe.' To be sure, if there is 'really' no such thing as wrong, then hurting people could not be called wrong. And even if right and wrong do really exist, one could still argue that hurting people is in itself morally neutral or indifferent—like, say, incinerating books, where the rightness or wrongness depends entirely on the context. While I personally lean toward a 'moral realism' position, which deems right and wrong to exist as real qualities or characteristics of human actions, I do not attempt here to support that position. Rather, I am taking it as a given that, absent justification, hurting people is wrong.

[11] A 'Humane Principle' that provides limited justification for inflicting human suffering as a way to avoid alternatives that are even less humane is presented at greater length in my earlier paper, *supra* note 3.

[12] I have specifically in mind here, by analogy, the objective standard of 'reasonable' care, *i.e.,* the care that would be exercised by a 'reasonable' person, in the law of negligence, which assumes that there is a legitimate objective that is sufficiently worthwhile to justify the risks of foreseeable harm that are inevitably entailed in every human action. *See, e.g.,* the classic United States v. Carroll Towing Co., 159 F.2d 169 (2d Cir. 1947); *cf* People v. Hall, 999 P.2d 207 (Colo. 2000) (criminal standard). *See generally* W. PAGE KEETON ET AL., PROSSER AND KEETON ON TORTS 173-217 (5th ed 1984); RESTATEMENT (SECOND) OF TORTS § 283 and comments (1977).

justification are not sufficiently shown, then the justification fails.[13] And if the justification fails, then the infliction of human suffering remains unjustified, and morally wrong.

Of course, empirical facts can never be shown beyond all doubt and, therefore, no fact-dependent justification can ever be absolutely certain. Nonetheless, people routinely act in the face of uncertainty, and it would be unreasonable (indeed, paralyzing) not to. But just because certainty can never be absolute, it does not follow that anything goes. The strength of a justification depends critically on the degree of certainty that there is about its underlying factual predicates.

What degree of certainty should be required to justify something as serious as the intentional infliction of human suffering? A plausible minimum standard might be what is sometimes referred to as *reasonable* certainty—the degree of certainty that 'ordinarily prudent' people require as a basis for action generally.[14] While ordinarily prudent people require widely varying degrees of certainty, depending on the tradeoffs of foreseeable harms and benefits, discounted by risk, this much seems safe to say: More certainty would normally be required for acts that have potentially grave consequences (such as serious human suffering) than is needed when the worst foreseeable outcome is trifling.

When it comes to something as morally suspect as deliberately causing human suffering, a reasonable person concerned to avoid wrongdoing would need relatively robust proof that free will exists before accepting it as a reasonable predicate for action. If the factual existence of free will has, for example, only weak and highly contested empirical confirmation, and if hurting people is presumptively wrong, then to inflict human suffering based on its supposed existence would be, by definition, unreasonable.[15]

To put this in a concrete context, consider the punishment imposed under the criminal law. The deliberate infliction of human suffering in this context—currently one of major activities of government—is frequently justified by the idea that offenders 'deserve; punishment because they freely *choose* to offend. As it is sometimes said, they bring it on themselves. But this justification for infliction presupposes the factual existence of free will. It ultimately stands or falls on the presupposed fact that free will exists. By definition, it would be unreasonable to punish a person based on her own free choices unless it were reasonably certain that she actually *had* a free choice. However, if we lack reasonable certainty about the very possibility of free choice, then all we can say with confidence is that the government engages in the naked infliction of human suffering.

---

[13] Analogously, in law a plea of self-defense fails if the defendant did not in fact have a reasonable belief that the use of force was necessary. *See, e.g.,* People v. Goetz, 497 N.E.2d 41 (N.Y. 1986). *Cf.* MODEL PENAL CODE §§3.04 and 3.09(3) (1962).

[14] Here again, I have specifically in mind, by analogy, the objective standard of 'reasonable' care, *i.e.,* the care that would be exercised by a 'reasonable' person, in the law of negligence. *See supra* note 12.

[15] That is to say, even if the infliction of particular suffering were endorsed by large numbers of 'ordinary' people who are generally 'prudent,' that would not in itself qualify such inflictions as 'ordinarily prudent.' In order for a reality check of this sort to be valid, it would necessary also to know that the ordinarily prudent people in question also genuinely accept (as opposed to merely say) that hurting people is presumptively wrong.

The urge to retribution runs strong in humankind.[16] People look hard for justifications to rationalize their appetites to act on this urge. The ideology of free will is pivotal in an important class of these justifications. If free will cannot be factually demonstrated, the moral implications would be particularly unpalatable: It would deprive us of our rationalizations. If it is wrong to deliberately hurt other human beings, and the freewill pretext for 'just deserts' turns out to be feckless, it would mean that it is morally wrong to hurt even wrongdoers.[17] It would mean that the offenders we punish are victims of the same chains of cause and effect that led to their offenses: "Before being causes they are effects."[18]

When faced with criminality we are morally obliged to do *something*, of course: To simply ignore those who commit harmful acts among us would pose unreasonable risks to everyone. Therefore, for pragmatic and utilitarian reasons, when inflictions are inseparable from reasonable programs of incapacitation, carefully calibrated deterrence or honest efforts to rehabilitate, they are mitigated and justified by their tendency (or, at least, purpose) to protect others from harm. But when human suffering is inflicted for such purposes as retribution, expressing social outrage or providing an 'appropriate response,' the inflictions are in a different category: Without a sound theory of just deserts, social programs that are deliberately designed to reach out and hurt human beings as a *goal*, and not as merely a deeply regretted side effect, would not enjoy the mitigating justification that they traditionally have claimed. They would be an evil that is unmitigated.

III.

Evidence For and Against the Existence of Free Will. Among the strongest evidence that free will exists is our own introspection, the seemingly unmistakable feeling we have that we make our own decisions. After all, we rarely if ever feel like there are external causes that 'dictate' what we do. On the contrary, the choices we make, even when powerfully influenced by one or more factors, seem basically *uncaused*—that is to say, free. Every day, at many points, we feel we have any of a number of different choices and, although some options may seem more attractive than others, we hardly ever feel like there is anything that literally limits our options to a single one. This introspection is probably the best evidence we have in favor of free will.

On the other hand, grave doubt is cast on the existence of free will by both (a) experimental evidence tending to exclude the possibility of *conscious* free will, which is the only kind we can introspectively know, and (b) the basic mechanics of natural selection, which make it extremely unlikely that evolution could ever produce a biological species possessing freedom of will. Experiments and the logic of evolution may not irrefutably prove that free will does *not* exist, but

---

[16] I have discussed biological (evolutionary) reasons for this in my earlier piece, *supra* note 3.

[17] *Cf.* P.F. STRAWSON, FREEDOM AND RESENTMENT (1962), urging, as an argument for free will, the alleged impossibility or, at any rate, profound undesirability of giving up the natural 'reactive attitudes' of disapproval and moral condemnation that people feel toward those who do wrong. However, in light of the almost certain biological origins of these reactive attitudes, *see supra* note 3, their existence is but weak evidence for free will. It is impossible to see how resentments, 'moral sentiments' or like emotions that evolved to serve the special needs of life on the Pleistocene veldt could provide useful instruction as to the nature of 'real ' right and wrong—not at least without, at some level, indulging the naturalistic fallacy that 'is' implies 'ought.' *See* HUME, TREATISE, *supra* note 8, at 469.

[18] Sénik, *supra* note 2, at 35; *Accord,* HUME, ENQUIRY, *supra* note 8, at 92 ("While we act, we are, at the same time, acted upon.").

they seem to raise enough doubt about its existence to prevent free will from being a matter of reasonable certainty—not sufficiently certain, at any rate to support something as morally suspect as the intentional infliction of human suffering.

a. Our introspective knowledge of free will. If freedom of will exists in fact, it is an existence we know little about. It is a bit like a black box: We have subjective impressions about the supposed *effects* of free will, but that is about it. No one has observed the underlying mechanism by which an independent 'will' can actually deflect human muscular movements and, hence, affect human behavior. What we do know is this: The brain and central nervous system have physical features, such as mass, extension in space, and microscopic structures that have various electro-chemical properties. The micro-components of this physical system operate as a complex, adaptive input-output device according to the usual universal physical laws. They take in raw perceptual information from the sense organs, they process it through a neural network, and they send out motor impulses to the muscles, which cause human behavior. As a device to control behavior, the brain's neuronal interconnections and synaptic firings are essentially electrochemical (computational) and therefore deterministic. There is no empirical evidence that any event ever occurs within the brain or central nervous system that does not have an ordinary physical cause.[19]

Unlike the brain and central nervous system, the human 'will' has no mass or extension in physical space, nor indeed any detectable physical existence or structure at all. It is, to put it bluntly, essentially immaterial or spiritual.[20] And while there is a well-substantiated and empirically

---

[19] The details of neuroscience, while a fascinating topic, do not play a necessary role in the case being made in this essay, which deliberately tries to remain austere in its factual predicates, invoking only the most salient and non-controverted findings of neuroscientific investigations, namely, that the brain (along with the central nervous system) operates as an input-output device, that takes in raw perceptual information from the sense organs, processes it through an elaborate neural network, and sends out motor impulses to the muscles, generating behavior—and, in the process, changes itself at the micro-structural level (especially synaptic strengths) in ways that can affect later motor responses (*i.e.,* it learns). Among the sources that I have consulted in forming my own understanding of the structure and workings of the brain, mind and central nervous system are: PAUL M. CHURCHLAND, THE ENGINE OR REASON, THE SEAT OF THE MIND (1996); IRA B. BLANK, INFORMATION IN THE BRAIN: A MOLECULAR PERSPECTIVE (1994); RICHARD M. RESTACK, RECEPTORS (1994); ANTONIO R. DAMASIO, DESCARTES' ERROR: EMOTION, REASON AND THE HUMAN BRAIN (1994); PATRICIA S. CHURCHLAND AND TERRENCE J. SEJNOWSKI, THE COMPUTATIONAL BRAIN (1992); OWEN FLANAGAN, CONSCIOUSNESS RECONSIDERED (1992); NICHOLAS HUMPHREY, A HISTORY OF THE MIND (1992); ARNOLD TREHUB, THE COGNATIVE BRAIN (1991); DANIEL C. DENNETT, CONSCIOUSNESS EXPLAINED (1991); PAUL M. CHURCHLAND, A NEUROCOMPUTATIONAL PERSPECTIVE (1989); ROGER PENROSE, THE EMPEROR'S NEW MIND: CONCERNING COMPUTERS, MINDS AND THE LAW OF PHYSICS (1989); DAVID H. HUBEL, EYE, BRAIN AND VISION (1988); JOHN H. HOLLAND ET AL., INDUCTION: PROCESSES OF INFERENCE, LEARNING AND DISCOVERY (1988); PATRICIA S. CHURCHLAND, NEUROPHILOSOPHY: TOWARD A UNIFIED SCIENCE OF THE MIND/BRAIN (1986); JEAN-PIERRE CHANGEUX, NEURONAL MAN: THE BIOLOGY OF THE MIND (1985).

[20] My own guess is that the will, as we experience it, is what might appropriately be called an 'emergent epiphenomenon.' That is to say, it is a phenomenon that depends on underlying physical events for its existence but which *emerges* out of those physical events and is seen by us something quite different from them (as, *e.g.,* a TV 'picture' emerges out of but is quite different from the luminous dots that comprise it, or the *meaning* of a printed page exists and is experienced by us as something quite distinct from the physical bits of ink on paper out of which it emerges). While it is easy enough to understand how human understanding can recognize or construct emergent epiphenomena as part of our normal conceptual powers, it is not so easy to see how the epiphenomena could turn around and modify the physical substrates out of which they 'emerge' and on which they depend. That is, in terms of the will, even if we can imagine how the physical brain can form and modify our 'free will' as we experience it,

corroborated explanation tracing the motions of our muscles and bodies in a continuous physical trail to electrochemical activity in the sense organs, central nervous system and brain, it remains a complete mystery how muscular contractions (and therefore, human conduct) could be either *initiated* or *deflected* by an non-material 'free' will. If free will has an effect on human behavior, it would have to be because, at some point in the generation of human movements, the 'will' causes or prevents the firing of synapses, thereby modifying the synaptic chain reaction that leads the muscles to contract. That is to say, to affect behavior the 'will' would somehow have to intervene and cause muscular contractions (actions) that are different from those that would have resulted from normal physical process. But how the 'will,' with no physical substance, could intervene in the electro-chemical synaptic chain has never been explained.

As long as it has never been explained, much less empirically proved, how a non-material entity such as 'free will' could intervene in physical processes to initiate or deflect human action, the perceived role of free will in human behavior remains essentially a matter of faith.[21] It is, on the other hand, possible to give a plausible account of how our strong introspective feelings of free will could be merely an *illusion*—albeit an enormously resilient illusion.[22]

The ways we think about many things are, of course, famously affected by illusions that diverge from empirical fact. We still say, for instance, that the sun rises in the east, not that the earth's rotation carries us in from the west. Even when we know the empirical facts of a matter, our more natural tendency is to view the world from our personal perspectives. The reason that free will seems so intuitively obvious is because most of the myriad factors that determine our choices, and the neuronal channels through which they exert their control, are ordinarily invisible to us. It is natural to suppose that what we never see does not exist.

But just as we have learned the truth about sunrises, we now know that our brains take in perceptual information from the world and generate responsive motor impulses within a vast realm of neuronal activity that we are not directly aware of—what is sometimes called, in common parlance, the 'subconscious.'[23] Even such everyday actions as walking and speaking require thousands of minute perceptions (*e.g.,* of position, balance, sounds, etc.) and responsive muscular adjustments. We could scarcely walk or talk if the needed neuronal processing occurred (or even registered) in our conscious awareness.[24] The countless needed synaptic firings take place deep in

---

how could our experience of free will turn around and have physical effects on the brain? It would be like being able to make the ink on a book page move around into new patterns according to the meaning that we want to see there.

[21] There is no intention here to impugn faith or belief. Indeed, my starting presumption that 'hurting people is wrong' is essentially a matter of faith or belief. What this essay does maintain, however, is that neither subjective belief nor faith can, in itself, justify the intentional infliction of human suffering. *See supra* Part II.

[22] As Descartes observed, "there is nothing we comprehend so evidently and perfectly." DESCARTES, PRINCIPLES OF PHILOSOPHY 1, § 41.

[23] It well-settled among neuroscientists that we are actually conscious of only a small fraction of the total electrochemiical activity among the neurons in the brain. On the relationship between conscious and unconscious levels of decision making, *see, e.g.,* Francis Crick and Christof Koch, *The Unconscious Homunculus,* in THOMAS METZINGER, ED., NEURAL CORRELATES OF CONSCIOUSNESS 103-10 (2000). For an extensive and insightful discussion of the role of non-conscious mental processing and the missteps that have occurred by overlooking its importance to considerations of what is "voluntary," *see* Denno, *supra* note 2, at 308-37.

[24] When talking of neuronal activity as being 'in' conscious awareness or 'in' consciousness, all that is meant is that the activity is accompanied by subjective sensations of perception, decision, recall, thought, daydream, or the like

the shadows without being consciously felt or otherwise sensed. And while we might like to hope that our more important decisions emerge into consciousness for advance review, the point is that the brain controls and directs most muscular movements (behavior) at the unconscious level.[25]

The fact that unconscious neuronal activity makes up a major fraction of what goes on in the human brain does not, of course, rule out the possibility that the will is free. It does, however, provide a plausible account for how we could have a strong sense of free will even if it is only of an illusion. When no causes are discernable, there seems to be no cause. Since the proximate determinants of muscular movements are hardly ever felt nor otherwise experienced, our behavioral choices seem non-determined, a result of free will. Hence, the illusion. However, knowing what we now do about the huge proportion of non-conscious activity in the brain, we can no longer confidently assert, based merely on subjective introspection of our *conscious* minds, that the human will is free. Like sunrises in the east, our sense of free will could just as easily be a surrogate for the truth, that is to say, an illusion.

Now there is nothing wrong with acting on illusions, provided they are harmless. Indeed as surrogates for truth, illusions can be very handy expedients. The illusion of the sunrise is a classic example. Closer to the topic at hand, there is, for example, the well-known illusion that we feel pain at the locations of our body where the impingements occur (*e.g.,* a pinprick is 'felt' in the fingertip, not in the brain). Similarly, our visual sense presents us with objects that appear to be 'out there' in front of our eyes when, in fact, all we actually see are neural images in the backs of our heads. Each of these illusions is a surrogate for truth, but they are highly useful even though they do not strictly accord with the actual fact.

Analogously, free will is a story that we tell to ascribe *authorship*—the idea that some of the events in the world can be ultimately attributed to particular human beings.[26] The concept of free will and its derivative, authorship, are intrinsic to how we think and express ourselves on questions

---

that fall within our own personal experience of the internal and external events of the world. It means, in other words, that the neuronal activity in question is not within that portion of overall neuronal activity as to which we are oblivious. The use of the word 'in' emphatically does not presuppose the existence of a separate conscious area of the brain, or anything of the kind. (No such area has ever been found.) Which specific mechanisms cause some neuronal activity to be accompanied by conscious experiences (such as representations of perceptions, decisions or other mental events) while most neuronal activity is not so accompanied is, as far as I know, unexplained. *See* prior footnote and DENNETT, *supra* note 19, at 33-39 (on the modern debunking the mind/body 'dualism' of Descartes).

The best explanation I have seen is that of Dennett, who seems allow (in his 'multiple drafts' theory) that definitive conscious experiences of brain activity (representations of perceptions, decisions, etc.) need not all come together as on the fly at all, and may only be experienced, if at all, in retrospect, as recollections—possibly well after the fact. *See* DENNETT, supra note 19, at 134-38; and *id.* at 125-26 and 169 ('no moment can count as the precise moment at which a conscious event happens''); *but cf. id.* at 137 (seeming to insist on need for 'rolling consciousness' in order to negotiate a car through traffic); and *id.* at 169 ('no moment can count as the precise moment at which a conscious event happens'') If indeed we do not routinely have an on-the-fly conscious experience of brain activity in consciousness, summoning recollections only when we need them, this would imply, among other things, that what is normally referred to as 'introspection' is, in reality, retrospection (*i.e.,* re-creation).

[25] That is, by means of organized electrochemical synaptic activity of which we have no conscious awareness. *See* citations *supra* note 23.

[26] Authorship can alternatively be ascribed on the basis of a person's 'character,' *i.e.*, the person's own distinct combination of virtues and vices, which characteristically generates certain behavioral tendencies, such as generosity or crime. *See infra* Part V for further discussion.

of right and wrong (expressions that are, of course, both effects and causes of other people's behavior). As surrogates for truth, both free will and authorship are obviously useful. Imagine how cumbersome it would be if we could not attribute events to individuals but, instead, could 'see' only the lengthy chains of causation that lay behind them. Without free will and authorship, how could we ever preconceive or bring about, in interactions with others, the behavioral effects that we so naturally achieve through ordinary talk about right and wrong? How could we teach children to become well-socialized adults? The language of morality (and the thinking as well) would become hopelessly entangled in mountains of causal detail. As members of a social species we have critical survival interests in the behavior of one another, interests that make metonymic ascriptions of free will and authorship a practical necessity.[27]

The narrative of free will has a darker side, however. It is a story that we tell to attribute blameworthiness to persons and their actions, to divide the innocent from the culpable and to transform cause into blame. In earlier eons, in a world before much reason, attributions of blame may have been virtually indispensable to survival, serving as a means to signal threats posed by others so that risky interactions could be anticipated and avoided. But under modern conditions this aspect of the narrative can be dangerously misleading, and we can draw unwarranted moral conclusions from it. And when we do, free will becomes a story that we tell to separate people out for serious maltreatment, to identify those who 'deserve' the infliction of suffering. And it is a story that we tell to relieve ourselves of responsibility for the sufferings we inflict, laying the blame on the victims, who 'bring it on themselves.' Free will is the foundation of moralistic aggression, invoked to justify hurt.

In sum, the introspective, subjective impression we have that our own wills are free can surely be counted as evidence, but it is evidence that is equivocal and weak. It is evidence that can be plausibly explained as the product of illusion. Without some other corroboration, therefore, the evidence of introspection does not have much weight, especially inasmuch as the other evidence we have tends to point the other way.

b. Three Experiments. As is well known, people do not always consciously experience events at the exact instant they happen. This may occur for a variety of reasons. We look into the night sky and see starlight emitted millions of years ago. An astronomer sees a meteor hit the moon more than a second after it happens. There is even a delay, measurable in hundreds of milliseconds, between the time that photons hit the retinas of our eyes and the ensuing nerve impulses register visually in our brains. With these kinds of delays in perception the physical explanations (*i.e.,* the need for travel time) are well understood. There are, however, other perception delays, some of which have a direct bearing on the reliability of our familiar introspective feeling of free will.

---

[27] As discussed *infra* in Appendix A, to say that the will is not free does not at all mean there is no real or meaningful morality, or that actions are neither 'right' nor 'wrong.'

So compelling are these attributions of authorship, as surrogates for truth, that some have actually come to believe that they are real and have elaborated theories of personal responsibility based upon 'control,' which they imply. *See* Douglas Husak, *Does Criminal Liability Require an Act?* in ANTHONY DUFF, ED., PHILOSOPHY AND THE CRIMINAL LAW 60-100 (1998). It is, however, a bit like imagining that the industrial robots on a modern shop floor actually have responsible "control" over the chucks of material that they deftly manipulate.

While we may think we consciously experience our own 'decisions' more or less instantly as they occur, experiments show that this is not the case.[28] To the contrary, after the brain's neuronal activity settles on a behavioral choice, there is a delay before conscious awareness that the choice has been made. These delays can amount to 300 milliseconds (1/3 second) or more, about the time of a blink of an eye, and they mean that choices can be acted upon, via motor impulses to the muscles, before conscious reflection even becomes possible. Stated another way, we go through life 'seeing' the decisions we make roughly 1/3 second (or more) after we make them. The presence of this gap between consciousness and actual neural decisions is sufficient to raise serious questions on the possibility of *conscious* free will, *i.e.,* free will as we usually understand it.

One set of experiments that show this delay concerns the so-called 'phi effect,' discovered by Gestalt psychologist Max Wertheimer in 1912.[29] The phi effect refers to the illusion of motion that makes moving pictures possible. It occurs, for example, when a series of images is flashed in quick succession, each slightly out of register with the previous, so the objects in the images seem to move smoothly from one location to another—as in movies and on television. In appearing to move *smoothly* from place to place, the depicted objects seem to occupy a continuous succession of locations in the space between their starting and ending points—including locations where none of the flashed images actually showed them. The brain fills in the gaps. Experiments investigating this 'phi effect' carefully controlled the presentations of successive flashes and measured observers' ability to react to them. What the experiments appeared to show was this: People are able to "see" and act upon a flashed light *before* they actually become consciously aware of the flash.[30] The brain can process inputs and make choices at the unconscious level.

Experiments conducted by Benjamin Libet demonstrated that this phenomenon of unconscious decision making is not limited to suddenly reactive or reflex-type actions.[31] Libet connected people to electroencephalographs and, as they gazed at a specially designed clock, told them to flick their right wrists 'whenever' they wanted to do so.[32] Their electrically measured neuronal 'decisions' to flick were compared with the times of the reported conscious awareness of the decisions. The key finding was that the experimental subjects did not become consciously aware that they had decided to flick their wrists until at least 1/3 of a second *after* the measured onset of readiness potential in their brains: 'Clearly,' Libet writes, 'the brain process … to prepare for this voluntary act began about 400 msec. before the appearance of the conscious will to act'; 'the initiation of the freely voluntary act appears to begin in the brain unconsciously, well before the

---

[28] For a more detailed description of these experiments, *see* John A. Humbach, *Doubting Free Will: Three Experiments* (rev. January 12, 2010), *available at* http://ssrn.com/abstract=1535480.

[29] Wertheimer, M., Experimentelle Studien über das Sehen von Bewegung, *Zeitschrift für Psychologie* 61: 161-265 (1912). I became first aware of these experiments in Daniel Dennett's excellent and far more extensive treatment of the subject in his book CONSCIOUSNESS EXPLAINED. DENNETT, *supra* note 19, at 114-15, 162-66, 167.

[30] For a more detailed description of the phi experiments, *see* Humbach, *supra* note 28, at 2-6.

[31] Benjamin Libet, *Do We Have Free Will?*, 6 J. CONSCIOUSNESS STUDIES 47, 51 (1999), available at http://m0134.fmg.uva.nl/publications_others/BLfreewill.pdf.. The experiments are also described in FLANAGAN, *supra* note 4, at 136-38; DENNETT, *supra* note 4, at 162-66; and Andrew E. Lelling, *Eliminative Materialism, Neuroscience and the Criminal Law,* 141 U.Pa. L. Rev. 1471, 1520-26 (1995).

[32] Libet, *supra* note 31, at 50.

person consciously knows he wants to act.'[33] In short, decisions come first, then the conscious awareness of them.

Perhaps the most dramatic of these experiments, were those done by William Grey Walter.[34] His experimental strategy was similar to Libet's except that, unknown to the people being tested, Grey Walter had hooked the brain measuring probes directly to an external device, a slide projector. The people were told to advance the slides on the projector by pushing a hand-button at will. They were startled to see the slides advance, as per their decisions, *before* they pushed the button and, even, before they were consciously aware they had made a decision.

This ability that people have to process sensory inputs and act on them unconsciously is, of course, of great practical importance. A tennis player returning a volley can assess the ball's trajectory, decide where to place it and even hit it before she is consciously aware that her brain has 'seen' the ball coming across the net. The whole story of the ball's path and destiny arrives in conscious awareness only after it is all over. Of course, if human beings did not have this ability, we could not play tennis, ride a bicycle, drive safely in traffic, dodge a flying object (or a punch), or do anything else where fast reaction times are crucial. The larger point is, however, that our brains have the ability to make 'choices' and implement them by sending motor impulses to muscles, before we are even consciously aware that the 'choices' have been made.

Although these experiments show a systematic delay in our becoming consciously aware of external events and internal decisions on them, people do not of course *feel like* these delays occur. Rather, we *feel* like we perceive events and choices the instant they happen. We are not consciously aware of the delay between decision and awareness, just as we are not aware of the inherent delays that are needed for retinal response, optic nerve transmission or visual processing. The brain automatically compensates for all these delays by "referring back" our memories in time; the picture we get is, for example, that the hammer hits the nail at the instant of actual impact, not hundreds of milliseconds later, when our brain finally reports the scenario as a matter of our conscious awareness. Ditto if the hammer is coming down on our thumb: Even though we don't consciously see it coming down in time to respond, these experiments (and real experience) tell us that the brain can 'see' it coming down and even unconsciously send instructions to the arm to pull the thumb away.

c. Discussion of the experimental evidence. The experiments just described provide affirmative evidence that, at the very least, our brains do not make 'voluntary' decisions in quite the way that introspection tells us. Their conclusion is that people do not have consciousness of their decisions until about 1/3 of a second after they make them. The brain can and does send out instructions (motor impulses) to the muscles before there is conscious awareness that a decision has been made. Rather than consciousness *controlling* the brain's decision process, consciousness sits by and, as it were, passively *observes* the decisions we make hundreds of milliseconds after our brains have already made them. Before our choices are accessible in conscious awareness, all the

---

[33] Libet, *supra* note 31, at 50 & 51.

[34] Described in Daniel C. Dennett and Marcel Kinsbourne, *Time and the Observer,* in NED BLOCK, OWEN FLANIGAN, ET AL., THE NATURE OF CONSCIOUSNESS: PHILOSOPHICAL DEBATES 141, 168 (1997). Apparently Grey Walter only delivered the results of his experiments orally and never in writing. *See* DENNETT, *supra* note 9, at 240 n.2.

work of making the decision has already been done. This evidence of delay should at the very least cast serious doubt on the degree to which we can trust our subjective introspections about the existence of free will as a basis for purposefully inflicting human suffering.

But what about those instances when people clearly do consciously ponder and reflect on choices before reaching a decision? Do the experiments only apply to snap decisions? The studies done by Libet and Grey Walter suggest that the experimental evidence is not limited to snap decisions since, in neither case, were the subjects responding to immediate external stimuli that were prompting them to act. Rather, they simply sat quietly and decided 'spontaneously,' from time to time, to flick their hands or push a button. Thus, the Libet and Grey Walter experiments imply that, even when a decision is preceded by extended contemplation, the eventual resolution of the deliberative process (and, presumably, of each step in the thinking) does not become accessible in consciousness until after the usual 1/3-second delay. If this understanding of the evidence is correct, then the brain's unconscious processing is always the ultimate gatekeeper in determining what we decide. Indeed, the very threshold decision to engage in deliberation at all—to weigh the pros and cons in a premeditated way—is one that has to be made by processes that are unconscious.[35] So even if a final choice is made consciously, the decision to consciously choose has to be made by unconscious processes before any question of choice can enter into conscious awareness at all.[36]

There is still another way that decisions in the brain's unconscious processes inevitably impinge on conscious decision making. The factors that go into a conscious weighing of options must somehow be selected, and the weights these factors carry must somehow be assigned. As we would normally experience this process, the factors that are relevant to impending decisions just sort of 'come to mind'—popping into our consciousness ready for review. Inevitably the work of initially selecting these factors has to be done through unconscious processes. After all, one could not possibly do a conscious canvassing of one's stock of accumulated knowledge to select the memories that ought to be included as factors for consideration. Since the choices we make are essentially at the mercy of the factors that go into them, this hidden pre-selection process is crucially important, if not determinative. But what that means is that all of our conscious deliberation must be heavily biased if not totally dominated by unconscious neuronal activity. Even if our conscious thinking were 'free,' we would still have little or no conscious control over which things we think about.

But what about *unconscious* free will? Although the huge unconscious portion of decision making may cast a large shadow on free will as we experience it (*i.e.,* as a conscious act of choice), it certainly does not eliminate the possibility of free will at the *un*conscious level. But to solve the freewill problem by positing an *un*conscious free will is not without problems of its own. For one thing, such a possibility would seem to be completely conjectural and ad hoc, not even able to claim support from the evidence of introspection, since we do not introspect our unconscious minds. And there are other disquieting aspects as well. After all, having a free will just does not seem the same if the reins are held by sub-surface daemons whom we do not quite know and cannot really control.

---

[35] That is, by means of organized electrochemical synaptic activity of which we have no conscious awareness. *See supra* note 23.

[36] And if not the choice to choose, then the choice to choose to choose, and so on. At some point, activity of the conscious mind has to be turned on, presumably by something produced in the hidden workings of the brain, and that fact alone is enough to keep the will from being free.

In sum, while the evidence of these experiments do not definitively exclude the possibility of free will, they do cast serious doubt on the degree to which we can trust our subjective introspections about how we made decisions to act. And if we cannot rely on our primary source of evidence that free will exists, it would seriously undermine the factual basis of the freewill justification for purposefully inflicting serious human suffering.

IV.

Free Will and Evolution—Biological traits and characteristics do not last very long if they significantly burden the survival prospects of the individuals that have them. If freedom of will imposes a competitive disadvantage on those that have it, then it never could have evolved under the pressures of natural selection. The trait for free will, if it ever arose, would have soon died out.[37]

It appears on analysis that biological individuals with free will would be at a substantial survival disadvantage.[38] To see why, let us first consider the advantages that would accrue to the opposite of free will, namely, being 'rule-bound.' The idea of rule-bound (which applies to animate and inanimate systems alike) is simply this: Any given combination of extant circumstances will always lead to the same subsequent combination of circumstances. In other words, a 'rule' in a rule-bound system is the pattern:

if X, then Y

where X is a given combination of extant circumstances and Y represents the combination of circumstances that *always* follows after X. In the behavioral context, rule-bound behavior is the antithesis of behavior controlled by a free will. Rule-bound actions are not 'uncaused' (as free choices are) but they are instead the inevitable physical results of a widening pyramid of antecedent causes and effects that reach back indefinitely. Rule-bound actions are those that result from rule-bound psychological processes.[39]

---

[37] Daniel Dennett has presented an argument as to how a capacity for free will could have evolved in humans. *See* DENNETT, *supra* note 9, But the 'free will' that he refers to seems to be only a free agency (*see infra* Part V.b.) combined with an extensive power to pre-reflect and pre-ruminate on choices but not an actual ability to jump off the rails of physical determinism. *See id* at 225-26 (seemingly acknowledging that his conception of free will is just a 'complicated snarl mechanistic causes,' but insisting that it does the same job as 'genuine' free will), and 296 (seeming to concede that it is a deterministic world). *But cf.* the decidedly more mystical turn that Dennett takes *id.* at 268 (stating that 'our reflections will actually help determine which trajectory our future holds').

[38] Technically it is genetic 'reproductive fitness' rather than individual survival per se that determines which heritable traits persist and spread in natural selection. However, as a rule of thumb, if you don't survive you don't reproduce, and in this respect having a free will would probably be a disaster.

[39] Being rule-bound in rather like having the legal doctrine of precedent apply to motions and changes of the physical world, including an individual's behavior. In philosophical discussions the quality of being rule-bound is conventionally referred to as 'determinism,' but for this context I prefer the more expressive term 'rule-bound.'

It needs to be stressed, perhaps, that being rule-bound emphatically does *not* mean the rules are innate, instinctual, genetically determined or the like. Instincts may work for insects and birds, but I would guess that more than 99.99% of the behavioral 'rules' followed by humans are *learned*, from experience. That is to say, the only 'genetic' aspect of the rules is that we are disposed, in dealing with new experiences, to draw lessons from analogous past experiences and act in accordance with those lessons.

By contrast, a 'freewill' system, *i.e.,* one that *not* is rule-bound, means one that is capable of transitioning from one factual state to the next according to the pattern:

if X, then whatever

To say that a system is not rule-bound means, simply, that various different possibilities may follow whenever X exists and it is *indeterminate* which of them will actually occur; there is no rule that allows one to know what to expect, even as a matter of statistical probability. Such a system, because it is not bound by rules, can be called 'anomalous.' It can also be called 'free.'

There is, however, a very practical problem with having an 'anomalous' mind, one that is capable of breaking free of behavioral rules and following the pattern 'if X, then whatever.' The problem is that such a mind would be a terrible guidance system for negotiating an organism through life in a world where survival depends on detecting and adapting to the regularities of nature.

Things do not 'just happen' in the world around us. The forces of nature are not random but manifest themselves according to rules. When we wake up each morning, we do not find that the things of the world have been recombined or rearranged, so that yesterday's learning has become obsolete. The accustomed patterns are still there. The responses we have learned are still generally dependable. The events of the life continue to move in a largely predicable succession, and it is rare to be truly surprised. The physical universe is *regular* in its motions and changes.[40]

As a result of this regularity, nature at all levels is, despite its huge complexity, replete with patterns, and the correlations among its occurrences produce 'covariations' that are continually repeated. Biological individuals are able in varying degrees to detect these covariations and, by later recalling them, to anticipate the future by observing their surroundings. The more often that a given covariation has appeared in the past, the more probably it will recur in the future. And the more effective an individual is at detecting the recurrent covariations and anticipating what happens next, the more successful the individual will be, on average, at avoiding the hazards and capturing the benefits that life constantly presents.[41]

---

[40] While events at the submicroscopic quantum level are probabilistic rather than determinant, there are even at that level probabilistic 'rules' that statistically predict events, albeit not specific events. There are, moreover, almost certainly also rules setting limits on the physical possibilities: Nobody believes (I think) that a boson could bloom into a rose. At any rate, the phenomena of quantum randomness do not have any evident role in determining either individual decisions or the course of biological evolution. *But cf.* [Kane &] PENROSE, *supra* note 19.

[41] Strictly speaking, only the universe as a whole could literally follow the rule 'if X, then Y.' The behavior of any system smaller than the universe itself will always be subject to the intervention of "exogenous" factors that might produce results other than those dictated by 'if X, then Y.' Stated differently, the transitions in sub-universe sized systems can deviate from the 'rules' if the trajectory of events is deflected by exogenous factors whose existence was omitted in defining 'X' in the first place. (Such an omission is, precisely, an omission of the 'exogenous') As a practical matter, however, the environments within which most organisms live, though decidedly smaller than the universe, are nevertheless systems that move through time in a largely rule-bound way—that is to say, covariations detected by organisms are sufficiently likely to repeat themselves, rather than randomly deviate, that it generally benefits the organism to learn from experience.

It is true, of course, that perceived covariations can be deceptive, that nature does not present itself to the perceptions of individuals in a totally rule-bound way. For even if nature as a whole is rule-bound, the picture of the world that we collect from perceptions is always incomplete. We can never know every detail of what goes on around us, and the details we miss can sometimes be crucial. Despite this persistent incompleteness of information, however, the pictures that nature presents are still regular enough that experience is a generally good guide to what the future will hold. The ground is always wet after a rain. The lion that chased you yesterday will likely chase you today—if you don't watch out. In a complex but generally rule-bound world, an individual can, *on average*, derive a comparative survival advantage by responding to its surroundings in ways that it continuously adapts to the rest of what goes on.

A rule-bound individual thus acts to maximize the potentials that life presents by routinely modifying and refining its behavioral dispositions in light of the covariations that it has detected in the past, so its behavior will be as well adapted as experience can make it.[42] The rule-bound individual has, in short, the ability to learn from experience and to act reliably in accordance with that learning.[43] The more that the individual is intelligent, reflective or rational, the better it will be able to understand and use its experience. But if the individual decides to outsmart and override the lessons of experience, responding 'freely' according to the pattern of 'if X, then whatever,' it would, on average, miss opportunities and fail to miss pitfalls that the regularities of nature continually present.

Consider, for example, what it would be like for an individual that has a free-will mind. Guided by reflection and thought, such an individual could override the constraints of the rules it has learned and make freewill choices to do things *other than* what its experience would dictate for the conditions it currently perceives. Suppose, specifically, that such an individual makes a freewill choice to eat some tasty berries despite the fact that same berries yesterday led to crippling stomach cramps. By choosing to eat the berries, by choosing to do *anything* other than what experience-based adaptations would dictate, the individual would by definition be doing something that experience says would be probably *mal*adaptive. In a world that is generally regular, doing this would be (on average) a very bad idea.

Now, to be sure, experience is not always a good guide. In given cases the choices made by a free will might actually be *more* conducive to success than those of a rule-bound mind. After all, free will does not mean irrational, and in some instances a choice that *rejects* the teaching of experience might be better adapted to survival than one that is dictated by experience-based rules. This could occur for instance when, due to information gaps in a given situation, the individual's own particular prior experiences have been misinterpreted or are misleading, or when circumstances

---

By the way, there is no assumption that, from now on, future covariations will be like past covariations. However, looking back it can be seen that, for the individuals that lived and evolved in the past, *their* future covariations as of any given moment did turn out to be largely like their past covariations. And this is all that is required to make the point that is being made here about evolution, adaptation and free will.

[42] Behavioral dispositions are, ultimately, a function of the synaptic strengths and connections in the individual's brain. *See infra* Part V.b.

[43] *See* TIMOTHY H. GOLDSMITH, THE BIOLOGICAL ROOTS OF HUMAN NATURE 85 (1991) (the ability "to continually adjust their behavior with experience").

are encountered that seem familiar but are in fact novel. Consider, for example, a deer that is raised by humans and comes to regard people as benign and caring. If such a deer, released into the wild, later encounters a hunter looking for a kill, it could be catastrophic for the deer to act towards the hunter in a rule-bound way. The deer's experience-based disposition to approach and trust humans would probably lead to its death. By contrast, a freewill deer would at least have a chance of survival—by ignoring its lessons of experience, and fleeing.

This argument, as far as it goes, has merit since an individual has no guarantee that apparent covariations detected in the past were genuine, as opposed to illusory. Nor is there any guarantee that a situation similar to one from the past does not contain undetected discrepancies that might make the best response very different from the one that worked before. To generalize, the rules that an individual acquires based on the lessons of experience can sometimes produce bad results. For these situations, using free will may provide an advantage.

But in the end this strategy would lose out to the law of averages. The problem is that a freewill individual would never have any way to know *ex ante* when it should follow the lessons of experience and when it should try to outguess them. But every time it tried to outguess experience, the odds would be against it in a generally rule-bound world. In such a world, the observed regularities in nature are more common than the irregularities. The probabilities are that the lessons of experience will be not misleading but predictive.[44] In short, from a survival standpoint, the behavioral choices of a freewill mind would be worse *on average* than those of an adaptive rule-bound mind.

Finally, it should be noted, there is no basis to suppose that superior intelligence, reasoning, reflection or communication (such as enjoyed by humans) would improve the odds for freewill decisions over those prescribed by experience. For sure, having more brainpower will, on average, give an individual a greater ability to detect covariations, organize them, weigh them and apply them to anticipate the future and to calibrate its behavior accordingly. It would, in short, permit individuals to derive all the more advantage from the teachings of experience in a generally rule-bound world. It would not, however, confer an ability to make 'free' choices that are, on average, better adapted to survival than the best that experience has to offer. Even today, when cultural factors can soften or even counteract biological forces, the world remains, on average, regular, as opposed to random or erratic. Using the tools of intelligence, culture and language we are obviously better able to understand and apply this regularity, which lies at the heart of science and technology. But there is no reason to think that these tools for interpreting experience could produce even better results if, in the course of using them, we were to disregard, in whole or in part, the experience that they exist to collect and apply. In sum, in a rule-bound world, despite our always imperfect perceptions of past events, there is no known way to improve on the behavioral responses that result from using our superior human intellect to evaluate, systematize, weigh, reflect upon and intelligently apply our accumulated experience-based knowledge of how the world works.[45]

---

[44] This statement assumes, of course, that the individuals in question have evolved perception apparatuses that allow them to reliably detect and memorialize associations of regularities and covariations in their environments. That is, the individual must be able to map the rule-bound world in a rule-bound way upon its rule-bound mind in order for its own internal adaptations to be beneficial to genetic survival.

[45] It is, by the way, by no means clear why the use of reasoning, reflection, meditation or the like in reaching a response should make the response any more free. Just because a neural-network calculation goes through a large

In sum, if a freewill individual ever emerged in the course of evolution it would, in a generally rule-bound world, labor under a serious survival disadvantage compared with its rule-bound competitors. Being disposed to sometimes reject what its learning would prescribe, overriding the lessons of past, its anomalous mind would lead it to choices that are, on average, worse than those of its rule-bound rivals. By ordinary evolutionary processes, these extra hazards and lost opportunities would mount up and translate into a comparative disadvantage in the competition to leave offspring in future generations. Thus, even if in the course of human evolution some truly free, anomalous minds emerged, the genes that made them possible would have quickly become extinct.

V.

Free Will, Virtue-Ethics and 'Compatibilism.' The freewill justification for punishment supposes that offenders deserve to suffer because they freely *chose* to do bad. An alternative basis for saying that suffering is deserved is that the offender *is* bad. Under this alternative basis, moral culpability is traced not to a free but faulty choice but, instead, to a flaw or 'fault' in the very nature of the person being punished.[46] On this view, it would be morally right to inflict suffering on a bad person (say, a pedophiliac) because harm emanates from what the person is—namely, someone who, by disposition or character,[47] spawns abhorrent behavior, like fruit from a poisonous tree.[48]

a. Virtue ethics as a substitute for 'free will.' The idea that people can deserve to suffer based on what they *are* goes back at least to Biblical times[49] and is reflected in common modes of expression and political stances that reveal how people often think about crime and punishment. When it is said, for instance, that society should lock up 'criminals,' there is a frequent implication that the persons to be locked up are of a different kind—not ordinary folks 'just like us' except they made some different choices, but dangerous beings that need to be removed. (After all, good people don't choose evil deeds, do they?[50]) This tendency to disaffiliate from and dehumanize those who

---

number of steps gives no basis for assuming that 'somewhere in there' it manages to shake free of the physical (electro-chemical, synaptic) rules of production that it ultimately rests on.

[46] *See, e.g.,* Powell v. Texas, 392 U.S. 513, 540 (1968) (Black, J. concurring) (explaining that, even though a compulsive alcoholic might get drunk due to something other than his own volition, the motivating force is nonetheless "'his' except in some special sense); Commonwealth v. Malone, 47 A.2d 445, 449 (Pa. 1946 (elevated culpability ("malice") based on "a *wicked disposition* evidenced by the intentional doing of an uncalled-for act in callous disregard of its likely harmful effects"; emph. added); and Bedder v. D.P.P., [1954] All E.R. 801 ("unusually excitable or pugnacious temperament") (Lord Simmonds, L.C.). *See also* HUME, ENQUIRY, *supra* note 8, at 89-95 (seeming to support character as the sole legitimate ground for punishment); Ekow Yankah, *Virtue's Domain* 19-29, *available at* http://ssrn.com/abstract=1265668.

[47] Note that, in the discussion that follows, the terms disposition and character are used more or less interchangeably, though this is contrary to the practice of some. While it is perhaps more conventional to reserve the word 'character' for those dispositions that are relatively deep-seated and durable, this distinction does not, as far as I can tell, make a moral difference or, apart from extreme cases (like mental illness), a legal difference.

[48] "[A] corrupt tree bringeth forth evil fruit [and cannot] bring forth good fruit. Every tree that bringeth not forth good fruit is hewn down, and cast into the fire. MATTHEW 7:17-20.

[49] *See supra* note 48.

[50] *Cf.* HUME, ENQUIRY, *supra* note 8, at 89-95; Yankah, *supra* note 46, 20 & 26 (describing how this thinking allows the creation of a 'criminal caste').

cross the line is particularly evident in the case of sexual crimes, which are (in common parlance and thinking) committed by 'perverts' and 'predators,' the latter being a term that literally means a kind of animal.[51] While one should not carry this sort of linguistic analysis too far, it does reveal an oft-encountered tendency to see those who offend as being in a category apart, with their crimes being viewed as marks of an essentially different and flawed personal character.

This way of thinking about fault is, of course, closely connected with the idea that people are endowed with differing mixes of virtues and vices, and it implies a kind of 'virtue ethics' basis for personal accountability: On such a basis, inflictions of human suffering can be morally defended on the ground that some people 'deserve' to suffer because of what they *are,* namely, criminals—playing essentially the same role in our lives as vermin, parasites or dangerous wild beasts.[52] After all, rooting out evil has got to be right, doesn't it? What is more, this virtue-ethics basis also makes it possible to see the fault for inflictions as something that *belongs to the sufferers* themselves. As such it provides a way, without resorting to factually dubious free will, to deflect responsibility for the suffering that governments inflict by insisting that the sufferers bring it on themselves. For even if persons of bad character cannot freely choose their own acts, even if they are prisoners of cause and effect, the fact remains that the fault for their harms lies within themselves. Whether an offender's fault consists of a free but flawed choice or of flawed patterns of neuronal synaptic strengths in his brain, the bottom line is the same: The fault is his own.[53]

To be sure, the modern criminal law does not regard 'bad' character alone as enough to justify punishment (just as a mere guilty 'will' is not enough).[54] The law also requires an 'act'[55] as a

---

[51] We would never use the term 'predator' to refer, for example, to the authorities that set up the sting operations to trap potential sex offenders. Yet, their endeavors are, objectively speaking, also rather analogous to the food acquisition ambuscades of carnivorous wildlife. Apart from terminology, it may be observed that the widespread demand for post-release registration of and residency restrictions on sex-crime convicts long after they have ceased to be ongoing 'offenders,' reveals this area as being one where, in particular, the individuals' crimes are viewed as proof of a persistently flawed character.

For an interesting analysis showing how the view that sex offenders are "non-human" is shaped and reinforced by the media's exaggerations and its distortions of the incidences and heterogeneity of offenders, see Robert Doyle, *Proposition 83, Framing and Public Attitudes toward Sex Offenders: An Application of Heuristic Models of Social Judgment, available at* http://ssrn.com/abstract=1444688

[52] Note that the kind of virtue ethics referred to in the text is not the benevolent sort of virtue ethics in which everybody is urged to pursue *eudaimonia*, but a rather more punitive sort, such as that of the Daniel Statman, *The Time to Punish and the Problem of Moral Luck*, 14 J. APPLIED PHIL. 129 (1997); Douglas Husak, *Does Criminal Liability Require an Act?* in ANTHONY DUFF, ED., PHILOSOPHY AND THE CRIMINAL LAW 60-100 (1998); Katherin A. Rogers, Retribution, *Forgiveness, and the Character Creation Theory of Punishment*, 33 SOCIAL THEORY AND PRACTICE 75-103 (2007) ;and Katherin A. Rogers, *Anselm on the Character Creation Theory of Punishment*, 4 THE ST. ANSELM JOURNAL (Spring 2007) , available at http://www.anselm.edu/library/saj/pdf/42rogers.pdf ("One deserves to suffer because one is wicked" [at 7] and "[P]unishment is justified because the one with bad character ought to suffer" [at 2]); or the Bible, *supra* note 48 ("Every tree that bringeth not forth good fruit is hewn down, and cast into the fire"). Today, explicitly punishing bad character is generally out of favor (and with good reason, in my view). But perhaps it has merely gone underground. For otherwise, what are we to make of such legal phenomena as the elevated penalties for 'hate crimes' and assaults involving genitalia, proactive programs to identify and incarcerate pedophiliacs, or punishing recidivists more harshly—all of which seem largely to amount to punishments of bad character. *See* Mirko Blagaric, Punishment and Sentencing 240-42 (2001) (recidivism enhancements). For an excellent discussion of the role of perceptions of virtue in justifying punishment, *see* Yankah, *supra* note 46, 19-29.

[53] *See supra* note 46.

[54] DRESSLER, *supra* note 4, at 85-88.

legal prerequisite for punishment. But this prerequisite for punishment should not be confused with the *justification* for punishment. If the law does not regard a harmful act alone as a sufficient ground to punish,[56] then a virtue-ethics justification can (in the same way as free will) be invoked to supply the moral cover that is needed.[57] That is, if the factual objections to free will keep it from being a reasonably certain justification for punishment, then the virtue ethics can substitute in its place. For whatever one may think of deliberately hurting people because of what they are, the virtue-ethics foundation for punishment at least justifies the infliction of human suffering based on something *real*—on the actual behavioral dispositions that are written into the physical neuronal features of the sufferer's brain.

b. Virtue ethics as the real 'compatibilism.' Compatibilism is the idea that free will can, if properly understood, be *compatible* with a purely physical universe, *i.e.,* determinism.[58] But the *compatibilist* version of free will turns out, when viewed from a different angle, to be essentially the same idea as the virtue-ethics foundation for justifying inflictions of human suffering.

At first glance, the basic idea of compatibilism might appear a contradiction: How could a rule-bound individual possibly have a free will? The two seem mutually exclusive. But the basic project of compatibilism is to find in human behavior a sufficient element or residue of freedom to provide a hook on which to hang moral accountability, and hence, punishment.[59] The compatibilists accept that human acts are determined (rule-bound) but argue that the apparent contradiction between determinism and moral accountability can be resolved if we simply make a small revision in the conception of free will: The freedom needed for accountability does not (they say) literally require that the *will* be free. It is enough for free will that the *person* be free—free to do what she wills.[60]

---

[55] MODEL PENAL CODE §2.01 (1962). But it does not necessarily have to be a harmful act—as in the case of the various inchoate crimes predicated on mistakes of fact. *See, e.g.,* MODEL PENAL CODE § 2.04(2) (1962) (mistake of fact), which come very close to explicitly punishing character by measuring culpability according to the harm the offender hoped to do, not according to the risk or harm that she was ever in an actual position to produce.

[56] That is, if the law does not punish for cause alone (as opposed to fault), so an act itself could not serve as justification for punishment (as it once may have, *see* Adekemi Odujirin, The Normative Basis of Fault in Criminal Law 29-39 (1998); Albert Levitt, The Origin of the Doctrine of Mens Rea, 17 Ill. L. Rev. 117, 120 (1922)).

[57] We may still insist on the prerequisite of an 'act,' of course, for at least two reasons: First, as evidence. What better proof of character than a revelatory act? Secondly, when an act causes harm or increases risk, it provides a marker to fix the proper *amount* of punishment—more harm calls for more suffering. Without the act, the case might seem to be one of 'no harm, no foul.' But if a bad act by itself would ordinarily not be enough to justify punishment, *e.g.,* if not done by choice, the offender's evident character fills the gap.

[58] A convenient and accessible précis of compatibilism and its various flavors can be found in the STANFORD ENCYCLOPEDIA OF PHILOSOPHY, *Compatibilism*, *supra* note 2..

[59] "For the most part, what philosophers working on this issue [of free will] have been hunting for, maybe not exclusively, but centrally, is a feature of agency that is necessary for persons to be morally responsible for their conduct." STANFORD ENCYCLOPEDIA OF PHILOSOPHY, *supra* note 2, at 3.1.

[60] *Cf.* "Free will is the unencumbered ability of an agent to do what she wants." STANFORD ENCYCLOPEDIA OF PHILOSOPHY, *Compatibilism, supra* note 2, at 3.1. "[N]o liberty can be inferred to the will, desire, or inclination, but the liberty of the man; which consisteth in this, that he finds no stop, in doing what he has the will, desire, or inclination to do." THOMAS HOBBES, LEVIATHAN , ch XXI (1651).

This revision in the conception of free will shifts the focus of the freewill inquiry away from the will and onto the person, treating freedom of the person as the essential equivalent of free will—at least the important part of free will. What such a move accomplishes is this: It allows moral accountability (and, therefore, inflictions of suffering) to be based on something that can be called 'free will' without running head-on into the objection that free will is based on a dubious and mystical (non-deterministic) factual premise.

There is, however, a problem with this compatibilist attempt to reconcile free will with determinism. The problem is that it confuses free will with *free agency*.[61] The two are different. When a person is a free agent, she is free to do what she wills, but that is not the same thing as having freedom of will.[62] To have a free *will* (as opposed to mere free agency), an individual's choices in a given situation would have to be free of constraint not only from her current external circumstances but also free from her internal neuronal state. To have a free will, the individual would have to be able to break free from the electro-chemical synaptic processes that otherwise would rule her muscles and direct what she does.

However, determinism (which the compatibilists accept) means, if nothing else, that individuals do not have the ability to jump the rails of physical cause and effect. That being the case, the only freedom that an individual has is free agency. And, on examination, it appears that when free agency is made the linchpin of moral accountability, the real basis for making people suffer turns out, once again, to be what they *are,* just as in virtue ethics. That is to say, what compatibilism actually proposes is, in the end, to punish people based on their character—or, at any rate, on their unchosen, neuron-based dispositions. In terms of justifying punishment, compatibilism and virtue-ethics rest on the same foundation.

To see why compatibilism and virtue-ethics are so similarly based, it is helpful to divide the vast web of causes of human behavior into two broad categories: situation and disposition.[63] Any human action at any given moment results from an interplay of the person's *situation* at that moment and her *disposition*. Stated another way, what a person does at any given moment depends on her immediately prior external circumstances acting upon the sum-total of her internal neuronal state.[64]

---

[61] Perhaps more precisely, it pretends to trade on the moral currency of free will but offers only free agency, as though it is essentially the same thing. *See, e.g.,* DENNETT, *supra* note 9, at 225. But, as explained in the text immediately *infra*, it is not. *Cf.* HUME, *supra* note 8, at 88.

[62] *I.e.,* A free agent is stuck, at any given moment, with the behavioral disposition she has and, while she is free to carry out the 'choices' that her disposition dictates, she is not free to choose, form or modify her disposition or otherwise affect what those 'choices' are.

[63] The terminology of situation and disposition is borrowed from the work of Edward E. Jones and Victor Harris [citations to be added], though the two terms are being used here to make a somewhat different point.

Recall, too, that in this discussion the terms disposition and character are being used more or less interchangeably. *See supra* note 47.

[64] What do dispositions consist of and where do they come from? The short physicalist answer would be that they reside in the brain. Specifically, they reside in the patterns of neuronal interconnections and synaptic strengths that have developed over the lifetime of the person in question. As an input-output device, the brain takes in raw perceptual information, processes it, and sends out motor impulses to the muscles, generating behavior. Because people have different patterns of neuronal interconnections and synaptic strengths, they have different dispositions and, in a given situation, respond with different behavior. And the reason they have different patterns of neuronal

Different individuals will of course normally react in different ways to a given situation. Suppose, for example, there is an unattended open convertible parked by the curb with a camera lying on the seat. Three people pass by and notice it. One passerby might see the camera but not have the least thought of taking it. The second passerby might weigh the pros and cons of stealing the camera, and then either take it or not. The third passerby might be so disposed that he grabs the camera without a second thought. These three passersby—all in the same situation—express three very different personal dispositions. In the given situation, the disposition (internal neuronal state) of each produces a different behavioral response.

With respect to some matters, most adults are very similar in disposition, so similar that their behavior seems to be virtually dictated by the situation. For example, people in burning automobiles almost always try to get out. Another classic example is found in the law of duress. When people receive commands under an immediate threat of serious physical injury, they virtually always comply. Once again, situation is essentially the only variable that determines what people do. To say that the situation is essentially the only determinative variable is just another way saying that, for the given situation, almost everybody has the same disposition, so almost everybody acts the same. We would say, in common parlance, there is 'no choice.'

In most situations, however, there does seem to be a 'choice,' and people show marked differences in how they behave, differences that can be attributed to their differing dispositions. For example, one coffee drinker may take a single lump of sugar while another takes zero or two. Less trivially, a customer who receives excess change at the supermarket checkout may or may not mention the error to the cashier. Some office workers see an unattended supply cabinet as invitation to steal; others do not. Some respond aggressively to a perceived slight while others react hardly at all. These differences in possible responses to similar situations might traditionally have been explained in terms of free will, but the less mystical explanation is simply that the people involved have differing dispositions.

In other words, in analyzing the sources of human behavior, disposition can fill the same determinative role that is often attributed to free will. However, free agency does *not* play that role and, indeed, it cannot. The reason it cannot play that role is because free agency is not a *source* of behavioral constraint but an *absence* of constraint. And obviously, an absence of constraint cannot 'determine' the way a person acts. When a situation is such that people are free agents who can act in different ways, it is their dispositions, not their free agency, that decides how they behave. But the fact that disposition and not free agency plays the determinative role means that, for the compatibilists, there has been an awkward realignment of causes.

Remember that the compatibilists accept determinism but they insist that there remains, in the springs of human behavior, a residue of freedom that can serve the moral function that has been traditionally served by free will (*i.e.,* provide a basis for condemning abhorred acts and justifying the infliction of suffering). The compatibilists find this residue in the freedom of the person, *i.e.,*

---

interconnections and synaptic strengths is because they have different personal histories of causes and effects that have laid down memories and otherwise molded the development of their central nervous systems. A person's disposition at any given time can (like the person himself) therefore be understood as an entity representing a culmination of particular causes and effects interacting down to the present moment.

free agency, which they claim is similar enough to free will to justify holding the person punishable for her acts.[65] However, as we have just seen, in a rule-bound (determinist) world it is not free agency that plays the crucial behavior-deciding role; it is the person's dispositions. Therefore, when bad behavior occurs it is not the person's free agency but her dispositions that should receive the blame. If a free agent appears to be faced with several possibilities and 'chooses' one of the bad ones, the resulting bad act is due her faulty disposition.

In short, when free will is redefined as free agency, the real ground for infliction turns out to be the complex of causal factors that have formed (determined) the free agent's internal neuronal state, *i.e.,* her disposition. The fault lies not in the wrongdoer's bad 'choice,' which was rule-bound, but in her disposition or 'character.' Just as in virtue ethics, the unchosen happenstance of having a harm-producing disposition is the ultimate foundation for moral condemnation.

And so, the compatibilist version of free will turns out to be to be essentially the same idea as the virtue-ethics foundation for justifying inflictions of human suffering, only viewed from a different angle. While both compatibilism and virtue ethics succeed is avoiding the dubious factual status of free will, both are nonetheless subject to the objection that they endeavor to justify blame and punishment based on a person's non-chosen, causally determined faulty disposition, ultimately the patterns of neuronal interconnections and synaptic strengths in the hapless offender's brain.[66]

So both compatibilism and virtue ethics raise the important question, can it be morally justified to purposefully inflict human suffering based on an offender's faulty disposition when her disposition was never chosen but was, instead, determined by forces outside anyone's ultimate control? In favor of the justification, it can be said that the offender's fault, whatever its causes, is a personal internal fault of her own, and that internal flaw alone may constitute a morally sufficient reason to inflict—much as people kill snakes, poison rats or step on spiders without serious qualms about the suffering that these efforts might occasion. One does not, moreover, have to look hard to find analogous moral sentiments aimed at human beings whose dispositions are such that they are regarded as threats or dangers: serial killers, child molesters, terrorists or even just muggers and thieves.

For sound biological reasons, inflicting suffering on dangerous or abhorrent people may feel justified because they 'deserve' to suffer as part of their very essence of their being.[67] At very least, as has already been observed, those who are a proven danger to others require utilitarian or

---

[65] *See, e.g.,* DENNETT, *supra* note 9, at 225.

[66] Indeed, bringing things around full circle, punishing people for what they *are* may be the whole point of the ideology of free will. After all, good people don't choose evil deed, do they? *Cf. supra* note 50 and accompanying text. For a compatibilist like David Hume, for example, or Daniel Dennett today, the free agency version of free will seems to fit perfectly with a general aim to punish the irredeemably bad. *Id.* and *supra* note 9, at 293 ('you really can't teach these old dogs new tricks, apparently').

[67] *Cf.* my discussion of the biological origins of the retributive urge in Humbach, *supra* note 3. Analogously, our modern urges to inflict suffering on dangerous and scary people, and the feelings of satisfaction derived from such inflictions, may simply have evolved as nature's way of motivating individuals to anticipate threats and to eliminate threatening elements from their own environment. I hasten to add, however, that being 'natural' does not necessarily make something right.

pragmatic intervention for the protection of all.[68] But the intentional infliction of human suffering on the ground of 'just deserts' is another thing altogether. Whether it can ever be morally justified to regard human beings in the ways we might regard venomous snakes, stinging insects or rabid dogs is, to say the least, somewhat dubious.[69]

VI.

A Few Last Words on Free Agency. While free agency may be illegitimately co-opted to defend inflictions of human suffering, it would be remiss to close without a nod to the immense moral importance that free agency has in its own proper sphere: After all, abundant legal protection of free agency is the essence of freedom under law. That is to say, free agency is certainly a variety of freedom 'worth wanting.'[70] And, lest there be misunderstanding, this point must be stressed.

But the freewill ideology contributes nothing to the importance of free agency. For modern human beings, free agency is vital as a value in itself. Whether or not there is freedom of will, a wide scope of free agency (or, more simply, 'freedom') is practically indispensable to happiness, personal satisfaction and a high quality of life. This means, among other things, that a government committed to freedom should endeavor to minimize, and certainly not add to the 'situation' factors in people's lives that could prevent or hinder them from doing whatever, given their dispositions, it is their will to do.

In so many ways, as we seek a high quality of life we are our dispositions. The human thirst for freedom is, at bottom, a thirst for the free agency. For it is only through free agency that can we actualize the desires that are embedded in our dispositions.

In short, the proper reason to recognize free agency is as part of a rationale to promote and extol human freedom. But it has no proper part in projects to inflict human suffering.

---

[68] In accordance with, in my view, the Humane Principle. *See supra* note 3.

[69] More recently, some compatibilist have argued that the sought-after residue of freedom to predicate moral responsibility can be found, if not in free agency alone, in the complex of abilities to reason, deliberate and communicate that are unique to human beings. *See, e.g.,* DENNETT, *supra* note 9. However, I cannot see how ability to reason, responsiveness to reasons and other such mentalistic hocus-pocus would somehow add enough choice to provide an independent predicate of moral responsibility, *i.e.,* enough to fix blame on the basis of something other than the offender's disposition to do wrong. For in the end all such mental maneuvers have to be rooted in and under the control of the individual's own particular happenstance of remembered experience, and the ultimate question is always the same. Was it possible (without indulging counterfactuals) for the offender to have used the reasoning, reasons and experience that the past had given him to produce an act that was different from the one he did, one that would not have been bad? And the answer is likewise the same: If he could have done so, then he would have. If he didn't, he couldn't, and to punish the offender for the bad act that he in fact produced is to punish for the past that created him.

What free agency plus responsiveness to reasons, *see* DENNETT, *supra* note 9, might help justify are certain utilitarian or pragmatic bases for punishment (especially for the purpose of deterrence). It would be, for example, seem to be a precondition for deterrence that either the recipient of the ill-treatment or some target audience of potential offenders would be responsive to such deterrents, as 'reasons.'

[70] *Cf.* title of DANIEL C. DENNETT, ELBOW ROOM: THE VARIETIES OF FREE WILL WORTH WANTING (1984).

VI.

Conclusion. The factual existence of free will is at very least contestable and there is, moreover, substantial evidence against it. This evidentiary situation is a serious moral concern. To the extent that unsupported factual assumptions support or motivate inflictions of human suffering (or lessen the inhibitions against them), the questionable beliefs and assertions are themselves a source of unjustified harm.

The ideology of free will is invoked to rationalize and justify the infliction of serious human suffering. But the existence of free will cannot be shown with reasonable certainty. On the contrary, there are both experimental evidence and evolutionary considerations that cast substantial doubt on the factual existence of free will. Accordingly, to punish people because of the 'choices' they make is a naked infliction of human suffering, barren of any reasonably certain justification.

Faith in mythic entities and forces are the right of everyone. But when mystical entities and forces are invoked to justify the deliberate infliction of serious human suffering, that is something else again. Human society has a long history of purporting to justify the infliction of suffering by pure acts of faith. Generations of witch trials, punishments for blasphemy, massacres of heretics and crusades against 'infidels' attest to the harm that unsupported belief can produce. We have happily moved past most of that, and free will remains one of the last redoubts of faith-based infliction. Until there is reasonably certain proof that free will exists, it cannot justify doing harm.

Appendix A

The Possibility of Morality in a World Without Free Will. If the will is not free, can there still be a meaningful place for ideas of morality? Can it make sense to call actions 'right' or 'wrong'? My answer is, of course, yes. Indeed, the foundational point of this essay is that hurting people is presumptively *wrong* and that the ideology of free will is an major engine of wrongdoing. Still, one might object, if every human action is pre-determined, like the firing of pistons in a machine, what sense would morality make? If we are all just playing out our lines in roles written for us billions of years ago, when the universe was formed, what is the sense of praising some actions as 'right' and denouncing others as 'wrong'? Isn't every human action just a morally neutral fragment of the cosmic clockwork drama?

The short answer is that there is no connection between what something is and the fact that it is preordained. Thus, the fact that actions may be all pre-ordained is no reason at all why actions cannot be either 'right' or 'wrong.' No one would say that a red wagon is not 'red' just because its color was deterministically ordained. Similarly, moral rightness and wrongness, as characteristics or qualities of actions, do not depend for their 'existence' on free wills any more than colors do.[71] So

[71] The word 'existence' is place in quotes because, of course, as Locke pointed out, colors are secondary qualities, dependant on the observer for what they look like, and the moral qualities of actions are probably also secondary qualities. As observer-dependant secondary qualities, the moral qualities of actions would (like colors) thus be epiphenomena which emerge from the primary physical qualities without having necessary correlations with any *particular* primary qualities. *See* JOHN LOCKE, AN ESSAY CONCERNING HUMAN UNDERSTANDING 110-19 (1690: A.D. WOOSLEY, ED 1964). *See supra* note 20 re epiphenomena and emergence.

there is no reason why actions cannot be morally right or wrong even if they are not freely decided.[72] Similarly, even if, lacking free will, people have 'no choice' but to discern the moral qualities of actions and engage in discourse about them, that would not mean that such discernment and discourse are futile or inconsequential. On the contrary, moral discernment and discourse would still serve as the very as mechanisms by which changes are put into motion in the ongoing world of cause and effect.

Ascribing authorship to actions,[73] discerning their rightness or wrongness and announcing what we have discerned during moral discourse are among the ways that people participate in the intertwined chains of causes and effects that are the grand progression of the universe. Personal authorship may be an illusion, along with our conceptions of morality, but it can nonetheless serve as a surrogate for truth, which we can effectively use as a rule-of-thumb for action. So even if our ideas of authorship and morality are not 'true' (just as it is not 'true' that the sun rises in the east), these surrogates for truth can and do inform our thinking and ordinary language on matters of right and wrong, giving us the words we use to cause effects. They allow us to sidestep the long and intricate causal chains that go back indefinitely, and to keep the catalogs of causes at manageable levels, by attributing human actions to the persons who do them. They allow us to focus our harm-avoidance energies on the more easily controlled links in the causal chains (*i.e.,* those embodied in people), and to teach our children to become well-socialized adults. In short, even if human behavior is entirely "mechanical," social interaction can still benefit[74] from rules of the road. Accordingly, humankind's gradual discovery and implementation of these beneficial[75] rules can be seen as part of the grand cosmic program. The uncovering and internalization of ever better moral judgments, themselves matters of cause and effect, can affect and improve human social behavior.

And in the course of that improvement, the extirpation of moralistic aggression would be a good next move.

I personally am inclined to believe that right and wrong are real qualities or characteristics of actions, or that they are, at least, as 'real' as colors or other such secondary qualities. While it is true that red things appear to be red only in our heads, and wrong acts likewise may appear to be wrong only in our heads, it is still quite possible both such qualities emerge from (and are truth-surrogates for) deeper phenomena or things-in-themselves that we may not ever be able to know directly. Just as the colors on a relief map show things that exist, though not in the actual *way* they exist, our reason-revealed conceptions of right and wrong may draw attention to 'real' distinctions but show them to us in a way that is other than how they 'really' are. Ultimately, however, like all lines of reasoning, moral reasoning must start somewhere on a foundation of assumed premises, and the assumed premises will always contain the conclusions. While this may mean that, as a search for ultimate truth, moral reasoning must always be in some sense futile, it is no more futile than the physical sciences, which are ultimately grounded in the same soil of faith.

[72] I do not mean here to confine 'actions' to external, objectively observable events, but I allow it to include the benign or 'culpable' brain states that (could) lead to external actions. Therefore, if one wishes to think in terms of guilty minds or choices, for example, I see no difficulty in attributing moral qualities or characteristics to brain states as well as the actions they can produce.

[73] *See supra* Part III.a.

[74] *I.e.,* 'benefit' in the eyes of its participants, as they are pre-determined are see it.

[75] *See* preceding footnote.