1-12-2010

Doubting Free Will: Three Experiments

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

Doubting Free Will: Three Experiments

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

At another place I have argued that hurting people is presumptively wrong and that, because this is so, any act to inflict human suffering needs to be justified.[1] Moreover, I have claimed, the justifications for infliction require evidence, and because infliction is presumptively wrong, the burden is on the inflictors to justify the infliction.[2]

This moral presumption against infliction of suffering has particular relevance in the field of criminal law because, as a system of retribution, the criminal justice system is institutionally committed to the infliction of human suffering. Simply put, suffering is inflicted to make people "pay" for the crimes that they commit. Of course, the criminal justice system often has other goals as well (*e.g.,* prevention of crime through deterrence, incapacitation or rehabilitation). Nonetheless, insofar as retribution remains a significant force shaping criminal justice policy, the intentional infliction of human suffering is the aim.

A core justification for the intentional infliction of human suffering on criminals is rooted in the idea of "free will," the notion that criminals freely choose to engage in criminal conduct. For if criminals did not choose to commit their crimes, if their acts were literally dictated by causes beyond their control, then criminals would seem no more to be morally "responsible" for their misdeeds than the victims who suffer the wrong. The law would routinely excuse a person who steals with a gun to his head or whose car is hit from behind and strikes a pedestrian.[3] And though the causal chain may be more complicated or obscure in the case of human conduct, there is no known physical reason (*i.e.,* non-mystical reason) why this same principle could not be extended to criminal conduct generally.

However, the mere suggestion that this principle should be extended criminal acts generally sounds insensitive and morally obtuse. So deep and powerful is our human appetite for retribution that the mind recoils at the thought that it could possibly be wrong to make criminals pay for their crimes. The fact remains, however, that millennia of inquiry and debate have still not produced definitive evidence that free will either does or does not exist. The evidence of "free will" needed to justify retributive infliction still has not emerged. On the contrary, recent developments in the study of the brain and its role in human behavior have cast serious doubt on the free will hypothesis. There is no empirical

[1] *See* my companion paper, *Free Will Ideology: Experiments, Evolution and Virtue Ethics* 1-3 (January 12, 2010), *available at* SSRN: http://ssrn.com/abstract=1428002. This paper, supplying and discussing additional details of certain experiments, is intended to supplement and support the companion paper. *Id.*

[2] *Id.* at 3-5.

[3] In the words of the Model Penal Code, such a person would not be considered to have committed a "voluntary act." MODEL PENAL CODE §2.01 (1962).

evidence that any event ever occurs within the brain or central nervous system that does not have an ordinary physical cause.[4]

This working paper describes three previously reported experiments that tend to cast doubt on free will. All deal with the phenomenon that, for a variety of reasons, people do not consciously experience events at the exact instant they occur. There is, for example, 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. This kind of delay in perception is, of course, well known and its physical explanation (*i.e.,* the need for travel time) is well understood. There are, however, other perception delays that may be more surprising.

The experiments described here concern the delay that intervenes between the time that we make a "decision" and the time that we actually know, or consciously experience, that the decision has been made. While we usually feel that we consciously experience decisions at the instant they are made, experiments show this is not the case. On the contrary, the experiments appear to show that people do not become consciously aware that they have made a behavioral choice until at least 300 milliseconds (1/3 second) or so *after* the choice has been made. In other words, we go through life 'seeing' the decisions we make roughly 1/3 of a second (or more) after we make them. The existence of these delays is sufficient to cast serious doubt on the possibility of conscious free will, *i.e.,* free will as we usually understand it.

The three experiments described here demonstrate this 1/3 second delay. They are the Phi Effect, Libet's experiments and the experiments of W. Grey Walter.[5]

1. The Phi Effect—The "phi effect" (discovered by the Gestalt psychologist Max Wertheimer in 1912[6]) refers to the illusion of motion that makes moving pictures possible. The phi effect occurs, for example, when a series of images is flashed before the eyes in quick succession, each slightly out of register with the previous. The result is that

[4] 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). It should be stressed, however, that the details of neuroscience, while a fascinating topic, do not play a necessary role in reaching the conclusions sought to be supported here or in the companion paper that this one supports. *Supra* note 1.

[5] 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 4, at 114-15, 162-66, 167. Other citations to reports of the experiments are along with the pertinent text discussion *infra*.

[6] Wertheimer, M., *Experimentelle Studien über das Sehen von Bewegung*, Zeitschrift für Psychologie 61: 161-265 (1912).

the objects depicted seem to move smoothly from one location to another, as they do in movies and on television. Moreover, in moving *smoothly* from place to place, the depicted objects appear to occupy a continuous succession of locations in the space between the starting and ending points—including locations where none of the actual images put them. This illusion occurs because brain fills in the gaps; instead of a series of jerking moves we see only one continuous transition. This is the "phi effect."

The phi-effect experiment of particular relevance here involves two small spots of light that are flashed close to the eyes in quick succession.[7] The lights are arranged so they are both within four degrees of visual angle. Typically, each light is flashed for 150 milliseconds, with a 50-millisecond interval of darkness between them. As might be expected from the phi effect, the observer does not see two discrete flashes but instead sees a small luminous "ball" that appears move swiftly from the first flash location to the second and then disappear.

Now a question that presents itself is this: Is the observer consciously aware that her brain initially 'sees' the two successive flashes *as flashes* (at the instants when her brain initially registers the processed nerve impulses from each)?[8] The answer is apparently not. At any rate, the observers in the experiments do not *remember* seeing either flash. All they remember seeing is a smoothly moving ball.

But the odd thing revealed by the phi-effect experiments is that, even though the observers apparently never recall seeing the two flashes *as flashes*, it seems that their *brains* can 'see' them, process them and react to them. Specifically, when the observers are told to immediately push a button if the first light is red, but not otherwise, they manage to perform the task. Apparently, the observer's brain (if not mind) can 'see' the first light come on and alter her behavior in response, before the second light comes on. Moreover, they do so with reaction times that indicate their brains are able to process the first light and generate a correct behavioral response whether or not the second light ever comes on. That is, their brains are able to 'see' the first flash, and regulate the observer's behavior in response to it, in less time than it takes for the second light to appear (*i.e.,* less than 200 milliseconds).

What the phi experiments also reveal is that the observer sees the first light in different ways (flash or moving ball), depending on whether the second light comes on. If the second light does not come on, the first light will appear as simply a single non-moving flash. If, on the other hand, the second light does come on, the observer never recalls seeing the first light as a flash, but only as a smoothly moving ball. What is more, the ball seems to start moving *at once*, roughly 200 milliseconds (1/5 second) before the second light comes on. If there are several possible locations for the second light to

---

[7] Apparently the first to do this experiment were Kolers and von Grünau. *See* Kolers, P.A. & von Grünau, M. , *Shape and Color in Apparent Motion*, 16 VISION RESEARCH 329-335 (1976). Analogous results, but using the tactile rather than visual sense, have been produced by the so-called 'cutaneous rabbit" experiments done by Frank Geldard and Carl Sherrick. *See* DENNETT, *supra* note 4, at 142-43.

[8] The phrase "the brain …'sees'" is a metaphorical shorthand. What it refers to is the point at which the impulses originating at the retina have arrived the brain and been processed to such an extent that a neuronal representation or "image" of the visual scene has been formed.

appear, the direction of the ball's motion appears to be toward the place where the second light will later come on. In short, the observer seems to 'know' in advance whether and where the second light will appear (so as to form the picture of the moving ball) as much as 200 milliseconds before the time of the second light.

An even more dramatic phi-effect experiment involves two flashes of different colors—red and green. Now as the two flashes are presented, the illusory moving ball appears to abruptly change color, from red to green, halfway between the locations where the actual flashes occur. Obviously, the brain does not have the information it needs to correctly change the color (produce a different colored illusion) until after the raw data from the second flash has been processed off the cerebral end of the optic nerve. Once again, the observer seems to 'know' in advance what is going to happen before the second event occurs.

Assuming that the observers in these experiments are not actually able to look into the future, what the phi-effect experiments show is there must be some delay between the time when a person's brain 'sees' something (and is able to act on it), and the time when the person becomes conscious of what she sees. The observers in the phi-effect experiments could not have been seeing the first light as a moving ball until *after* their the second light came on, yet their brains were able to adjust their behavior based on the specifics of initial flash *before* the second light came on. That is, long before[9] the observers saw the first light in their conscious awareness, their brains had already 'seen' it and could have acted on it and moved on to something else. Conscious awareness of the two successive flashes was delayed until after both events had occurred (and was modified, as well), but there was no such delay in the brain's ability to discern what it 'saw' and to control behavior accordingly.[10]

The phi-effect experiments mean at least this: If the observers in the experiments had in fact ever consciously seen the two lights flash on and off, the memory of it must have been obliterated—replaced by a false memory of a moving ball. Another explanation is, more simply, that people do not see *anything* in conscious awareness until a many milliseconds after the scene has been 'seen' (unconsciously) and possibly even acted on by their brains. Though other evidence[11] seems to favor the 'delay' hypothesis over the false-memory hypothesis, there is probably not for present purposes much practical difference between them. If there are conscious goings-on inside our heads that

[9] That is to say, 'long before' in terms of brain processing time. Other experiments, described below, indicate that the delay in something under 1/3 second.

[10] Note that, in addition to the delay I am focusing on here, there is also some delay (mentioned earlier) in the retina's response, in the travel down the optic nerve and in the brain's visual processing to form the cerebral "image"—*viz.* the pattern of synaptic discharges that first reconstitutes the photonic impressions on the retina as a visual 'object' in the brain. These delays are not, however, the ones that we are talking about now. The question here is whether concerns the delay between the time that the visual processing regions of the brain have produced their visual object, of which one will be consciously aware, and the time at which one actually *is* consciously aware of that visual object. The evidence of the phi-experiments is that there is such a delay.

[11] *See* discussion of Libet and Grey Walter experiments, *infra.*

we never find out about because our memory of them is routinely obliterated within milliseconds, they are pretty much indistinguishable from 'unconscious' mental events.[12]

One last point: Even if there is a systematic delay in our becoming consciously aware of images and decisions on them (and, by implication, other thoughts and sensations in our brains), we nonetheless *feel like* we see things as happening the instant they occur. That is to say, we are not consciously aware of the delay. Nor, for that matter, are we consciously 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—so 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 to our conscious awareness.[13] Ditto if the hammer is coming down on our thumb: Even though we don't consciously see it coming down in time to respond, the phi-effect experiments indicate that the brain can 'see' it coming down and even unconsciously send instructions to the arm to pull the thumb away.

This ability to process inputs and act on them unconsciously has, of course, great practical importance in a number of ways. It means we are far more effective than we otherwise would be where quick reactions are required. A tennis player returning a volley can assess the ball's trajectory, decide where to place it and even hit it before (in real time) she is consciously aware that her brain has already 'seen' the ball coming across the net. The whole story of the ball's path and destiny arrives in conscious awareness only

[12] The most satisfying explanation for me seen is something on the order Dennett's so-called 'multiple drafts' model. It supposes, as I understand it, that various provisional representations ('drafts') of external encounters go flashing around the brain's neural network as the perceptual information comes in, until 'some threshold of activation' is exceeded 'over the whole cortex or large parts thereof,' at which point a resultant subjective experience arrives in consciousness. *See* Dennett, supra note 4, at 168 and 101-38; *see also* Stephen M. Fleming et al., *When the Brain Changes Its Mind: Flexibility of Action Selection in Instructed and Free Choices*, Cerebral Cortex 7, doi:10.1093/cercor/bhn252 (Feb. 11, 2009) (describing the decision between internally generated choices as being prompted "perhaps simply because of randomness in neural firing"), *available at* http://www.fil.ion.ucl.ac.uk/~sfleming/Fleming_CerCor2009.pdf; Jeffrey M. Beck et al., *Probabilistic Population Codes for Bayesian Decision Making, 60* Neuron 1142-52 (2008) (describing experiments done by co-author Alexandre Pouget), *available at* http://www.bcs.rochester.edu/people/alex/pub/articles/BecketalNeuron08.pdf (on efficacy of neural networks in resolving probabilistic problems through the use of mathematical attractors). Moreover, I think that Dennett's model would allow (and I tend to believe) that definitive or 'final' conscious representations of perceptions and decisions do not necessary ever have to 'come together' and arrive in consciousness at all as we navigate our way through life. Instead of routinely experiencing a running final-edit representation of the world generated (with delay) 'on the fly,' we would normally have conscious experience of a final-edit version, if at all, only in *retrospect*, in the form of recollections—possibly well after the fact. *See* Dennett, supra note 4, 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 internal and external events, 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.

[13] See discussion of Libet and Grey Walter experiments, *infra*.

after it's all over. Of course, if we 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 response 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.

2. Libet's experiments—Experiments conducted by Benjamin Libet likewise show that 'the initiation of the freely voluntary act appears to begin in the brain unconsciously, well before the person consciously knows he wants to act.'[14] In Libet's experiments, people were told to gaze at a specially designed clock and flick their right wrists 'whenever' they wanted to do so.[15] They were also told to pay close attention to the clock and note the exact instant when they first felt the urge, desire, or decision to flick. The people in the experiment were connected to electroencephalographs, which showed the exact time when their brains became ready send out motor impulses to their hands (the so-called 'readiness potential'). When the brain is *ready* to send out motor impulses to the body, that means, in effect, that the decision to act has been made.[16]

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.'[17] (The actual wrist flick came about 200 milliseconds later still.[18]) In short, decisions come first, then the conscious awareness of them.

To anyone who has jammed on the brakes in a car when a person or object suddenly appears in the road just ahead, the idea that decisions to act can occur 'unconsciously' is perfectly familiar. Nevertheless, the implications of Libet's experiment put a particular focus on the question: Is conscious awareness a necessary part of decisionmaking at all? After all, the actions of Libet's subjects were not sudden, reactive or unexpected, but were in pursuance of instructions to sit and flick their wrists whenever they wanted to do so. They had to make up their minds. Some of the subjects reported that they 'roughly' preplanned their flicks, but the time results were the same with or without such preplanning.[19]

The fact that consciousness of a decision appears significantly *after* the neuronal decision has 'a fundamental impact on how we could view free will.'[20] Apparently, however, Libet himself (as a freewill aficionado) does not interpret his experiments as

---

[14] 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).
[15] Libet, *supra* note 14, at 50.
[16] Flanagan, *supra* note 4, at 136.
[17] Libet, *supra* note 14, at 50.
[18] *Id.*
[19] Libet, *supra* note 14, at 50 & 53.
[20] Libet, *supra* note 14, at 49.

necessarily excluding a role for conscious choice.[21] He points out that the people in his studies said they felt like they were still able to intervene and *veto* the actual wrist-flicking once they became aware of their decisions to flick. And, Libet reports, such a veto was physically possible: The timing of the events would leave a 1/10 second window of opportunity between when the decision appeared consciously and the impulses went out to the muscles, and during this 1/10 second interval a countermand could be interposed.[22] As he put it, 'the conscious function [has] the possibility of stopping or vetoing the final progress of the volitional process, so that no actual muscle action ensues.'[23] This being so, Libet posits a version of free will that would work like this: 'We may view the unconscious initiatives for voluntary actions as "bubbling up" in the brain. The conscious-will then selects which of these initiatives may go forward to an action or which ones to veto and abort, with no act appearing.'[24] This power of final conscious control, Libet contends, is the locus of action for free will.

The interpretation by Libet of his own data is, however, a bit of a mystery. If it takes 350 milliseconds for the initial neuronal 'decision' to appear in consciousness and another 150 or so for it to start down the pipe to the muscles, why does Libet assume that a veto decision can spring to life and have its full effect almost *instantly*? It would seem more consistent with his experimental results to suppose that the conscious 'vetoes' reported by his subjects were, like the initial 'conscious decisions,' also 350 millisecond post-cursors of the actual operative decisions that they reflected. In other words, it seems more plausible (from the data) that *both* the decisions to flick *and* to veto were made unconsciously.

Libet, however, does not like the idea that the vetoes are decided unconsciously. To meet the objection that vetoes need (like the initial decisions) to be generated in unconscious processes, he points out that the factors affecting conscious veto decisions would not need to have *separate* preconscious processing; the factors entering into the conscious veto decision could be already processed as part of the package of factors that were preconsciously pulled together in generating the initial decision to act. From the freewill standpoint, however, there is a serious problem with this solution: First, it essentially concedes that the factors going into the initial decision, pro and con, are selected and assembled unconsciously. Then comes the opening for conscious free will, after initial decision and pro-and-con factors appear in consciousness. At that point the person has 1/10 second to make a conscious 'go-no go' judgment.[25] But on what basis is that judgment to be made? Based on the factors selected and assembled by the unconscious? By way of some random process? A quantum event? Libet argues hopefully that 'the *conscious decision to veto* could still be made without direct specification for that decision by the preceding unconscious processes.'[26] Maybe. But

[21] *See* Libet, *supra* note 14, at 51-53 and FLANAGAN, *supra* note 4, at 136. See the further very illuminating discussion of this in Lelling, *supra* note 14, at 1523-26.
[22] Libet, *supra* note 14, at 51.
[23] Libet, *supra* note 14, at 51
[24] Libet, *supra* note 14, at 54.
[25] 'The possibility is not excluded that factors, on which the decision to veto (control) is based, do develop by unconscious processes that precede the veto.' Libet, *supra* note 14, at 53.
[26] *Id.* (emph. in original).

relegating free will to 1/10 second snap decisions hardly provides a very convincing ground for imposing guilt and personal responsibility

If, on the other hand, we interpret the vetoes consistently with the other experimental data, we are led to the plausible conclusion that *both* the decisions to flick *and* to veto were made unconsciously. Consciousness then reflects but does not deflect the physical course of brain activity or motor impulses leading to behavior.[27]

3. W. Grey Walter's experiments—Perhaps the most dramatic experiments were those done by William Grey Walter.[28] He placed his subjects facing a screen on which photographic slides were projected. The subjects were told to use a hand-button to advance the slides at will. As in Libet's experiments, the subjects were connected to devices that measured the readiness potential in the area of the brain thought to subserve hand movement. As such, the devices could reveal when the 'decision' was made in the brain, and this decision-time could be compared with the (later) time when the subject pushed the hand-button. However, unknown to the subjects, Grey Walter had hooked the brain measuring devices directly to the slide projector. As a result, it was the readiness potential (decision) in their brains, not the hand-button, that actually advanced the slides. The subjects were startled to see the slides advance, as per their decisions, before they were even consciously aware that they had *made* the decisions.

Conclusion

These experiments do not definitely exclude the possibility of free will. They do, however, provide affirmative evidence that our brains do not consciously make decisions in quite the way that introspection tells us. This evidence should at very least cast serious doubt on the degree to which we can trust our subjective introspections about the existence of free will. And this doubt throws into question the factual basis of the freewill justification for purposefully inflicting serious human suffering.

* * * * *

For a further discussion of the implications of these experiments for the freewill justification of punishment, *see* my companion paper, *Free Will Ideology: Experiments, Evolution and Virtue Ethics* 1-3 (January 12, 2010), *available at* SSRN: http://ssrn.com/abstract=1428002.

---

[27] See also the generally concordant views of Andrew Lelling, at Lelling, *supra* note 14, at 1523-26. There have also been other explanations for Libet's results, notably Dennett's. *See* Daniel C. Dennett, Freedom Evolves 227-42 (2003). But Dennett's explanation tends to be significantly more intricate than the straightforward 'delayed consciousness' explanation and, contrary to the counsel of Occam's Razor, to 'multiply entities' that do not themselves have independent experimental verification (*e.g.,* 'free will takes time'). *Id.*

[28] 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 27, at 240 n.2.