

# Drive On Through: Sam Harris's Free Will

by Ian Kluge

Whenever he encountered a poorly made argument, one of my old philosophy professors used to say, "We could drive trucks through that one." What he meant was that an argument was so carelessly constructed that there were huge logical holes in it, problems like non-sequiturs, equivocation, insufficient evidence and correlative fallacies. As I read Sam Harris's *Free Will* (2012) I could hear Rueben Bonney behind me, intoning, "We could drive big trucks through that one."

Harris's goal in *Free Will* is simple: To persuade us that:

Free will is an illusion. Our wills are simply not of our own making. Thoughts and intentions emerge from background causes of which we are unaware and over which we exert no conscious control . . . Either our wills are determined by prior causes and we are not responsible for them, or they are the product of chance and we are not responsible for them. (p. 4)

He adds that:

The popular conception of free will seems to rest on two assumptions: (1) that each of us could have acted differently than we did in the past and (2) that we are the conscious source of our thoughts and actions in the present. (p. 6)

Finally, to make sure everything is crystal clear he states:

The intention to do one thing and not another does not originate in consciousness – rather, it appears in consciousness, as does any thought or impulse that might oppose it. (p.8)

He also claims that "to actually have free will:

You would need to be aware of all the factors that determine your thoughts and actions, and you would need to have complete control over these factors. But there is a paradox here that vitiates the very notion of freedom – for what would influence the influences? (p. 14)

In other words, Harris believes that free will requires an infinite regress of effective causes and since such a regress is impossible, free will is impossible:

Let us examine his arguments.

Consciousness is central to Harris's argument because the concept of free will and moral responsibility require that our conscious decisions be effective in controlling our actions. In other

words, consciousness and conscious choice have agency and we are responsible for our decisions and acts. That, of course, is what most people mean by free will and that is precisely what Harris seeks to refute by arguing that consciousness is powerless to do anything but to reflect or report what unconscious brain processes have decided. Consciousness is not “the source of our thoughts and actions in the present” (p.6). The problem with his view is obvious: while it may ‘solve’ the free will problem, it does so at the expense of raising an equally devastating problem. If consciousness is so powerless, what is its role? Why do we even have it if the brain can make decisions and take actions without consciousness? Increasingly sophisticated robots do quite well without consciousness.

This, of course, is the problem that has always dogged epiphenomenalism, which is the technical name of Harris’s position: mental phenomena are simply the by-products of physiological processes. Our awareness is an artifact of processes in which it has no part. What epiphenomenalism cannot explain – even in principle – is the very existence of consciousness and the subjective experience of exercising our will. If we don’t need consciousness for anything including decision making and acting, what survival advantage does it confer, especially if does nothing but give us delayed recognition of decisions made deep in the brain? If hidden processes in the brain are making all the decisions, is there any need for ‘us’ to consciously know about them? If not, why do we?

Thus, Harris’s argument is such that to eliminate free will, he also has to eliminate consciousness by reducing it to impotence and by leaving it without any function in human life. It is not even clear from the book whether Harris is aware of this result.

We could, of course, argue that consciousness and the sense of free will are biological spandrels, i.e. accidental by-products of other evolutionary developments in our brains. One of the problems with this response is that the whole subject of ‘spandrels’ is bogged down in a definitional debate, i.e. it is not entirely clear what is a spandrel and what isn’t. Worse, all examples of spandrels, pendentives, corbels and squinches do actually serve a function, i.e. they are necessary for something. We cannot avoid their function – and the function of consciousness is precisely what epiphenomenalism cannot explain. Until we understand it better, it is irrational to eliminate something that is universally experienced.

Another problem with Harris’s argument is that his claims run completely contrary to virtually universal human experience. Few people - except a few off their meds who actually feel like puppets on invisible strings – experience their conscious decision and willing as ineffective. It is easy to say they are deceived just as people were once deceived about the sun going around the earth. However, in this case, we are not talking about an externally, physically observable and measurable phenomenon but of people’s subjective, internal experience of themselves from minute to minute. To say that all humans have been deceived about their most personal and intimate aspect of their lives strains credibility. We cannot categorically say it is impossible, but we can say that such an astounding claim requires astounding evidence. Harris doesn’t even come close.

Harris thinks he has hard evidence against free will in the famous experiments of Benjamin

Libet. To sum them up: Libet's experiments (1983) seemed to show that a person's brain unconsciously decides that a person will move or not move his finger before there was conscious awareness of the decision to move. The person is aware of his decision to move his finger only after his unconscious brain processes had already activated themselves and reported to consciousness. In other words, the activation of the brain occurs unconsciously before the subject is conscious of having made any decision. Harris claims this is clear proof that the conscious will is unnecessary in making and carrying out a decision because the brain processes are wholly sufficient to the task.

However, Libet's experiment does not even remotely prove what Libet and Harris say it does. In the first place, Libet's experiment itself is not 'clean' or straight-forward. The test subject has to observe a clock to report when s/he becomes aware of their intention and decision to press the button. When we are dealing in milliseconds, the additional cerebral processing required for becoming aware of our decision, observing the clock and registering our awareness by actually pressing a button inescapably complicates and distorts the results of the whole experiment. Indeed, until there is an instant, i.e. timeless transition from our first awareness of our decision to checking the clock to actually pushing the button – inevitably a time-consuming process – this experiment cannot actually prove that the brain activity recorded on the EEG actually precedes our awareness of our intention. Without definitive proof on this point, their conclusions do not necessarily follow from their information, i.e. they commit a non-sequitur. Evidence is lacking and it is not clear that such evidence is even obtainable when all processes take time.

Libet's experiment is 'tainted' in another way. Libet has previously told his subjects that they are to press the button as soon as they are aware of their decision and after they have looked at a clock. But this undermines the experiment itself. The readiness potential could be connected to three 'events' – becoming aware of having made a decision, checking the clock and pressing the button. Are Libet and Harris suggesting that only the last of these has a readiness potential? That checking the clock, interpreting what it says, associating it with a sudden awareness of a decision do not require their own readiness potential? This will not work if we assert that everything that happens in the mind is part of a physiological process; there cannot be any mysterious processes that is exempt from this need for a physiological correlate.

Finally, the experiment is 'tainted' because giving his subjects instructions beforehand to press the button is nothing less than telling them to get into the readiness-potential mode before they even start! Anticipation and possibly imagined actions would already put the subjects' brains in the readiness mode – or even just prime them for the readiness mode – and we would expect that this mode precedes any conscious awareness of a supposed 'decision.' Of course, it is difficult to prepare the subjects for the experiment without 'tainting' it in this way – but that only illustrates why we cannot take this experiment at face value.

But there is another serious problem. How does Libet know that the brain activity his EEG machine reports is "a decision" and/or an "intention" to press the button? Electrical read-outs of brain activity do not provide clues of their specific content.

No one can take an EEG read-out and determine the subject is reading Hamlet Act 1, Scene 1, Line 32. Without knowing that specific content, we cannot definitively connect the unconscious

brain-readiness potential to the act of pushing the button. Since we cannot make that connection, then Libet's claim is simply an assumption for which he has no evidence. Here, too, we find a non-sequitur. Even worse, Libet's and Harris's reasoning is a classic illustration of the fallacy *post hoc ergo propter hoc* – before this, therefore, because of this. If something happens before something else, the first must be the cause: there was lightning just before my dog died, therefore, the lightning killed my dog. Just because there was some brain activity before I was aware of my decision to press the button, noting the time and then actually pressing the button, does not necessarily mean the brain activity was the cause of my decision. As Hume could tell them, a sequential pattern does not allow us to assume causality. This is one of Harris's correlative fallacies.

Even if – for argument's sake – Libet's experiment is valid for rather mechanical decisions to press a button, does that also mean it is also valid for more complex intellectual tasks? Imagine a cashier returns \$5.00 too much change, what should you do? What could you do? What would your friends and relatives say? And why? Think of all the implications. Or, if we are reading Heidegger's *Being and Time* and we are trying to understand and decide the validity of two key concepts, "ready-to-hand" and "present-at-hand." Is there any reason to think this would be as simple as pressing a button? Or, we hear a phrase like "heart of glass" (Blondie) and try to decide what it and all its allusions mean. Or trying to decipher a mysterious expression on our loved one's face. What about scientific and philosophical thinking? The idea that we should simply accept that high order mental processes are essentially no different than button pushing maneuvers requires more evidence and better logical argumentation than either Libet or Harris provide.

Harris tries to avoid the logical consequences of determinism by suggesting it really doesn't matter.

The case I am building against free will does not depend upon philosophical materialism (the assumption that reality is, at bottom purely physical . . . even if the human mind were made of soul-stuff, nothing about my argument would change. The unconscious operation of a soul would grant you no more freedom than the unconscious physiology of your brain does . . . if you don't know what your soul is going to do next, you are not in control. (p. 12)

This is "nonsense on stilts" as the British philosopher Jeremy Bentham once said. Harris's case is entirely dependent on ontological materialism. This is evident in his assumption that "soul-stuff" is not significantly different from the physical brain stuff and, therefore, also has a conscious-unconscious structure and all the associated problems. It is hard to think of a better example of 'spiritual materialism'; the soul in his view is simply a diaphanous duplicate of the brain. Logically, of course, this is a category mistake, i.e. treating one kind of thing as though it were of another kind, as in ravens and writing desks.

For Harris's this claim to make any sense, Harris should be able to show how his assumption can be tested scientifically. How can we test the functions of something that is not material, i.e. the soul, by the scientific method which is designed to test only physical data? How can we measure something that is not physical? How does he intend to test to see whether the soul is divided into two parts? And why would he assume that something non-physical would have the same

structure and attributes as something physical? It's like basing our understanding of horseshoes on our thoughts about sunflowers. This category mistake makes it clear that Harris has never given serious thought to the concept of a soul and is simply spinning words. Consequently, there is no reason to accept his claim that belief in the soul results in the same determinism as his Laplacean materialism. Furthermore, he makes the same mistake with willpower, claiming that "Willpower is itself a biological phenomenon" (p. 38). Once again, a round of questioning begins. How would/could you measure it – even in principle? What units of measurement? What are its physical correlates if it has any? One is reminded of the utilitarian's plan to measure pleasure in "hedons" and pain in "dolors."

At this juncture one thing is clear: neither Libet nor Harris have provided any proof that our decisions are made unconsciously, i.e. that our sense of making conscious decisions is delusional. Once this is understood, we can safely reject Harris's claim that:

The popular conception of free will seems to rest on two assumptions: (1) that each of us could have acted differently than we did in the past and (2) that we are the conscious source of our thoughts and actions in the present. (p. 6; emphasis added)

(1) does, indeed, represent the "popular conception" but, as we have already seen, nothing in Libet's experiment disproves it. There is also a hidden ambiguity in Harris's denial of (1). Does (1) say that we could not have acted differently in the past if we had different information or does it mean that we could not have acted differently if everything were the exactly the same? To say we could not have acted differently if we had different information, is nonsensical by Harris's own standards since different causal factors would be involved and lead to a different result. If we mean that we could have not acted differently even if everything were exactly the same, then we have a truism. Two identical situations lead to the same result. In most people's understanding (1) means we would have acted differently with different information (e.g. life experience) which is obvious in any causal system. Harris denies this (p. 6). The problem with denying (2), i.e. the belief that we are conscious agents only works if we accept Libet's experiments as the final word. There is no reason to do so. "The popular conception of free will" is safe from Libet and Harris.

Harris also claims that we need to know and have control of "all the factors that determine your thoughts and actions" (p. 14) Why? Why do I need to know the bio-chemical processes in my stomach to choose porridge for breakfast? To make my choice I need to know only two things – whether I am hungry and whether porridge suits me this morning. Indeed, my choice might be constrained by the fact that if I cook porridge I will be late for work but my choice will not be constrained by my lack of complete knowledge of the digestive process or the effect of oatmeal on the brain. Once we realize that this requirement for total knowledge is a red herring, it is self-evident the falsity of Harris's claim that "there is a paradox here that vitiates the very notion of freedom – for what would influence the influences?" (p. 14). He is attempting to disprove conscious agency by showing how it leads to an infinite regress. But this does not work. There is no reason why a numerous causes cannot come to a culmination and require a decision – porridge or pancakes – which we can consciously make. Nor is there any reason why we cannot

weigh all the evidence for and against Kant's categorical imperative and then decide whether we believe it is true or untrue, likely or unlikely, useful in some cases or not. Of course, there is no action in the latter case, but there is no reason to think free will is limited to action.

There is another aspect to what Harris sees as the problem of not knowing all the factors involved in making a choice. In *Free Will* he asks:

How can we be "free" as conscious agents if everything that we consciously intend is caused by events in our brain that we do not intend and of which we are entirely unaware? We can't. To say that "my brain" decided to think or act in a particular way, whether consciously or not, and that this is the basis of my freedom, is to ignore the very source of our belief in free will: the feeling of conscious agency. People feel that they are the authors of their thoughts and actions and that is the only reason why there seems to be a problem of free will worth talking about. (p. 26)

Harris also wants to challenge the "feeling of conscious agency," which, in his view, is the origin of the "problem of free will." To eliminate free will it is necessary to eliminate "conscious agency" and with it, a conscious agent, i.e. a 'self,' 'ego' or soul. Sooner or later, even exploration of free will arrives at this core question: 'Is the feeling of "conscious agency" a delusion' or does this feeling point to the existence of something real?

Usually, "the feeling of agency" and of a conscious agent is connected with the idea of a 'self' or 'soul,' i.e. something that is non-material and, therefore, outside the control of physical cause and effect. If the existence – or even the possible existence – of such a non-material entity could be demonstrated, Libet's experiment and Harris's arguments would be severely, perhaps fatally, weakened. We maintain that at the very least, the possibility of such an agent can be rationally demonstrated.

Imagine that you open a book in a foreign language. No amount of physical analysis will provide a clue about what the book means because the meaning of the book is not physical and intrinsically reducible to physical things. The same is also true of a book in a language you do understand. Physically, there are only material marks on a page. As you read the page, your brain synapses fire electrical blips. And here is the key: in regards to their physical existence, the print marks and the brain-blips are the same.

They can be fully analyzed by physical means for their chemical or bio-electrical content contents but contain no hint of their actual meaning. The brain itself as a physical organ has no sense of meaning. However, this leads to a serious problem: who or what is to 'understand' the meaning of the text?

If we go to yet another 'reader' of the book, a machine that scans the brain-blips, the same problem repeats itself: brain-blips cannot tell us the content meaning of brain-blips. An infinite regress has begun – telling us, thereby, that this method of identifying an entity that understands meaning will not work. By process of elimination, the entity that understands the meaning of the text cannot be material. It cannot be a physical thing – which is exactly what the common belief in a soul, self or ego asserts. At the very least, the rational possibility of a non-physical entity which understands meaning has been established. This undermines the logical foundations of

Harris's argument and the "popular conception" of a conscious agent that makes decisions has been strengthened.

The Baha'i Writings call these non-physical 'things' – meanings, souls, self etc. – "intellectual realities" (Abdu'l-Baha, *Some Answered Questions*, p. 83). They are not sensible realities but they affect us nonetheless because we ourselves are souls which are non-material entities capable of comprehending meanings. Among these "mental realities" (ibid.) are love (not physical lust), "the human spirit" (ibid.), and "the power of the intellect" (ibid.). They are capable of acting in or causing action in the physical world. For example, consider the following two text-messages: (1) "We won the lottery! \$500M. Hurry home!" and (2) "Don't love you anymore. Bye. P.S, took your dog." Both can cause enormous trains of physical reactions – and yet your cell-phone remains silent: because it does not understand meaning. We know that meaning can cause worldly action. But how? How can a non-physical entity cause action in a physical one?

This is the old 'mind-body' problem, which Libet and Harris attempt to solve by reducing mind to body. Other philosophers accept a mind-body dualism. However, I believe the Baha'i Writings have at least the outlines of a rational solution. We shall try to explain this position in broad outline form. The Baha'i answer is sketched in the following statement:

Some think that the body is the substance and exists by itself, and that the spirit is accidental and depends upon the substance of the body, although, on the contrary, the rational soul is the substance, and the body depends upon it. If the accident -- that is to say, the body -- be destroyed, the substance, the spirit, remains. (Abdu'l-Baha, *Some Answered Questions*, p. 239)

Let us first understand what Abdu'l-Baha asserts. The technical terminology he uses is from Aristotle whose method of analyzing reality is found throughout the Baha'i Writings. (See Ian Kluge, "The Aristotelian Substratum of the Baha'i Writings", *Lights of Irfan*, Vol. 4, 2003 at [http://bahai-library.com/kluge\\_aristotelean\\_lights4](http://bahai-library.com/kluge_aristotelean_lights4)) To say that "the body is the substance and exists by itself" means that the body has its own unique identity which is independent of anything else. For example, my whippet Athena is a substance; she has her own identity and even if I died, she would go on. Furthermore, every substance has qualities or attributes, Athena is a black and white – but her color is accidental, i.e. she would still be a dog-substance if her color were gray or brindle. Here is the key: the attributes or qualities are the expressions of the substance, i.e. they are the way the substance manifests or appears in the material world. What Athena actualizes are the potential attributes of her physical substance.

Abdu'l-Baha reverses this argument: he asserts that the "rational soul is the substance" and body is accidental, i.e. the body is one way the soul expresses or manifests itself in the material world. The body is an attribute of the soul or mind. In other words, the relationship between mind and body is in principle the same as the relationship between a substance and its attributes, e.g. between Athena and her colors. However, whereas in the materialist argument accepted by Libet and Harris, the physical body is the source of the mind and its controller, in Abdu'l-Baha's argument, the soul or mind is the source and the controller. The two arguments are mirror images of each other.

There are, of course, other questions to deal with on this issue but for now it must suffice that we

have shown that contrary to Libet and Harris, the existence of non-physical realities is rationally possible, and consequently, Abdu'l-Baha's solution to the mind-soul/body problem remains a logically viable option. Let us briefly return to Harris's Free Will.

Free Will is a dangerous book insofar as it provides superficial, reductionist answers to profound questions of human nature. Those who accept Harris's arguments will find their thinking is shallower for doing so. Nonetheless, it is a book that should be read – if for no other reason than to learn from its mistakes, some of which we have pointed out in this review. It is an excellent illustrative summary of scientism at work, i.e. the attitude that we can solve questions about human nature and morality in a laboratory, and ignore the fact that the scientific method is not equipped to deal with such issues.

ueCopyright © Ian Kluge. All rights reserved.