*Freewill* by Sam Harris – An Absurd Being Commentary

So, for this article, I’m assuming that you have read Sam Harris’ book, *Freewill*. If you haven’t, it’s very short, more essay than book, and well worth a read because it raises some interesting points that any proponent of freewill needs to address sooner or later. In lieu of this, you could read my previous article which briefly outlines what I took to be his main ideas.

Somewhat surprisingly, I agree with much of what Harris says… if we assume determinism to be true; specifically, what he has to say about fatalism, quantum indeterminacy, compatibilism and moral responsibility. All of the above are often given as reasons for resisting determinism and Harris, quite correctly in my opinion, rejects them in this capacity.

The fatalist argues that if determinism is true, then we have no reason to do anything because whatever is going to happen will happen anyway, *no matter what I do*. This is quite clearly mistaken. Whatever is going to happen will happen but *only as the inevitable result of your actions* in a great chain of cause and effect. If the way the universe has unfolded since the Big Bang is such that you are going to be a movie star, then a whole host of *other* things have also been determined prior to that. Perhaps you will take an acting course, try to meet certain directors, audition for a certain role, etc. But if, instead of doing these things, you just sit at home, Christopher Nolan isn’t going to suddenly turn up on your doorstep and offer you the lead in his next movie. Determinism doesn’t say that certain things are *destined to happen*; only that whatever happens *had to* happen. If you are determined to be a movie star then you are also determined to take that acting course, meet that director, and do all of those other things that lead to this. Staying at home and waiting for Christopher Nolan isn’t an option, although it *seems like* it is.

The disconnect here arises if, when you think about your *determined* (not *destined*) future, you ask, “Why should I do anything then?” *as if you had a choice in the matter*. If determinism is true, you don’t have a real choice over *whether* to do anything. In a fully determined universe, such a question is absolutely meaningless. This is why asking it creates confusion.

Quantum indeterminacy is no help either. Even if the foundations of the universe do turn out to be probabilistic, this won’t tip the freewill scales one way or the other. What quantum indeterminacy says is that events (at the level of the quantum) are truly random (although in such a way that the outcomes are governed by laws of probability). This means that the future cannot be predicted with complete precision, even in theory. However, determinism doesn’t require the future to be predictable. All it says is that effects follow inevitably from causes and all causes are physical. Even if those effects are probabilistic in nature, they still follow just as inevitably and can’t be influenced by any free, uncaused agency.

Moreover, quantum indeterminacy doesn’t do anything for freewill because random variations in physical outcomes still doesn’t give a ‘free-willing’ consciousness any space to operate. Physical causes still produce physical effects, all in line with physical laws of cause and effect, just now with a little uncertainty thrown into the mix.

I’m going to come back to compatibilism and instead address moral responsibility next. This is an important issue because it is frequently cited as a reason for rejecting determinism. It objects to determinism by saying that if we don’t have freewill, then we can’t hold people morally responsible for their actions. Now, this isn’t an argument. At best it’s special pleading. For this to be an argument then you would also have to make a case for moral responsibility being real. Spoiler alert: saying, “Moral responsibility must be real or else we have no reason to be good” won’t cut it. Perhaps there is no reason to be good. Maybe that’s just the way the world is. “But then society will fall apart.” Maybe so. Tough. “But then what about the law?” What about it? Pointing out that the consequences of a theory are highly undesirable is no argument against it. This isn’t religion. You can’t argue that a thing must be true just because the alternative makes you uneasy.

Harris agrees that determinism doesn’t give us moral responsibility in the fullest sense of the word we usually expect, i.e. with each of us being the absolute, free and uncaused agent of our actions, but he argues we don’t need this. It is enough to hold you responsible for your actions as the body/brain mechanism you are.

Typically people worry that this creates problems regarding justifying punishment. Not so. We take steps to deal with dogs that are dangerous (from locking them up to putting them down) even though no one attributes freewill to them. Why could we not do the same with human beings who, on this account, also lack freewill? The fact that a criminal lacks *absolute* moral responsibility doesn’t mean we no longer have grounds to punish him or her.

Some things do change with a deterministic outlook though. Ironically however, they change for the better. If determinism holds, we no longer have reason to tarnish people as somehow being guilty ‘in the core of their being’ or ‘in their soul’. In other words, attributing blame becomes less *personal*. This removes our justification for hatred and revenge. We can no more hate a person who wrongs us than we can hate a tornado for destroying our house. Both are just following inexorable laws of cause and effect. Naturally, this fosters compassion, even for those who harm us (didn’t a famous carpenter say something about this once upon a time?), and a healthy dose of humility when we turn this on ourselves and realise that our own successes aren’t our own, (in the sense that *I*, as an absolute, free agent, deserve full credit for them) but rather, we were just lucky to have the life, parents, upbringing, brain, personality, temperament, etc. that we do/did.

Which brings me to compatibilism. Harris takes issue with most versions of compatibilism because they simply redefine freewill to the point where it is just no longer worth having. Most compatibilists accept that we are fully causally determined (there is nothing going on in the brain except physical processes obeying the law of cause and effect) but argue that freewill doesn’t require us to be absolutely free, uncaused agents. Essentially, they are making the same argument that Harris makes for moral responsibility. Harris re-interprets moral responsibility so that the “you” who is morally responsible is not some uncaused agent; rather “you” are nothing more than the sum of a whole bunch of purely physical causes starting with the parents you were born to and ending with the last book you read. This certainly requires the shifting of some traditional philosophical foundations but it doesn’t reduce “morality”, in the sense of the good/bad distinction of acts, to meaninglessness. On the other hand, redefining freewill so that the “you” who is free is not an uncaused agent but a physical body/brain conglomerate subject to causality, makes a mockery of the whole notion of freedom. If we’re free in any way that matters, we cannot be causally determined. You just can’t have it both ways.

Now, having acknowledged that determinism is not under threat from some of the popular arguments against it and even offers some benefits by eliminating the justifications we might have for hatred and revenge while at the same time promoting compassion, I am going to pull the rug out from underneath everything I have said and make the argument that we do have freewill.

The first of Harris’ assumptions about freewill says that we could have behaved differently than we did in the past; i.e. if we rewind time so *everything* (including the physical state of my brain) is identical to the way it was when a particular decision was made, I could make a different decision than I did the first time around. Harris rejects this; in effect saying that history could not have been other than it was.

He relies on materialism to do this[[1]](#footnote-1), arguing that we (including our brains) are physical organisms in a physical universe and all physical matter is beholden to the same physical laws, including causality. Now, if we were ourselves non-conscious machines merely studying humans through external scientific methods, this might be a more compelling argument. There would be fewer reasons to suspect that humans were anything more interesting than more complex and advanced versions of other basic physical, non-conscious things (the word ‘non-conscious’ wouldn’t even mean anything). But we aren’t limited to this external view. We also know ourselves from the inside, enjoying privileged access to an inner world of conscious activity which *it is like something to experience*.

There are three prongs to the argument I want to outline here to reject materialism. The first is thoughts themselves. Thoughts are such a basic and intrinsic element in our lives that we seldom stop to think about them (no pun intended). If I ask you to think about a cup of coffee, an image pops into your mind. Now where is this image? In your head? Of course not. If someone dissects your brain, no matter how hard they look they will never find a coffee cup hiding down there, no matter how many quantum “microtubules” they sift through. Now, if it doesn’t exist anywhere in space, how can it possibly be physical? This is not to say the coffee cup isn’t produced by a physical organ, it very clearly is, and yet it itself is categorically and absolutely *not physical*. Few materialists that I am aware of have bitten the bullet on this and declared that thoughts aren’t real, but that is what their position dictates. If physical matter is all there is, thoughts (being non-physical) are categorically impossible. Determinists refute their own theory every time they have a thought about it.

The second thing I want to discuss is self-reflective consciousness, that awareness we each have of ourselves. Imagine I am a man. Well, you don’t need to imagine that; I *am* a man. However, I am not a pure and complete man ‘thing’, *through and through*. What I mean by this is not that I am part man, part woman, but that I am not *just* a man, solid and complete; rather, I am consciousness *of* being a man. This is not just a semantic trick. If I were a man in the same way that my table is a table (the same way that all physical matter is what it is with no possibility of ever being anything else), I could never be an awareness *of* myself as *anything*; I would just *be* it.

Finally, that old chestnut, the hard problem of consciousness; the subjective, intangible, experiential ‘feeling’ that accompanies all experience; the notion that ‘it is like something’ to see a red hat and this feeling is different from ‘what it is like’ to see a blue hat. These experiential ‘sensations’ are called *qualia*. From here the argument is essentially the same as the one I employed regarding self-reflective consciousness. There is nothing it is like to be a table (or any other clump of matter) because matter (atoms, quarks, electrons, etc.), as any good physicist will tell you, just isn’t the kind of thing which experiences qualia. Given that materialism, the worldview Harris is leaning on here, asserts that we (including our brains) are nothing more than matter, how is it possible for my experience of red to be like anything at all?

You could try and argue that the above are just illusions, as indeed our intuition about freewill is purported to be, but it is one thing to claim we think we have freewill when we don’t and another to claim the same about thoughts (about other things or the self) and subjective, first person experience.

Now although I normally take a fairly dim view on people who claim it takes more faith to be an atheist than a believer, in this case, in the face of so much non-physical evidence, it certainly seems as if Harris, along with other materialists, is caught in a faith position. We have thoughts, self-awareness, and enjoy a subjective, experiential aspect to our lives… but because he dares not violate the reductive doctrine of science that all is matter and all matter is as unconscious as the clump of matter that is my table, then… well, we must deny what is right in front of our noses because materialism is too important a doctrine to come under fire.

It is true that I have no idea *how* non-physical consciousness/thoughts/experiences could arise in a physical world but I am not prepared to deny their (surely obvious) reality just because they don’t fit in with a materialist, scientific worldview where they uncomfortably stick out as an unresolved (and unresolvable, I suspect) anomaly.

Now, you will recall when I was refuting compatibilism I said that if we’re free in any way that matters, we cannot be causally determined. This section has gotten us halfway there. I have argued that there are aspects of ourselves (thoughts, self-reflective awareness, and subjective experience) which are non-physical. I haven’t disputed that these things all depend on physical activity in the brain though (which, being physical, will be causal) so it still seems as if I haven’t escaped causality. The final nail in this coffin will have to wait until later.

At any rate something is definitely wrong with the materialist stance. Essentially, I’ve argued that it is inadequate to fully describe human reality and, since we are part of the universe, cannot be a complete account of that universe.

I want to turn now to address Harris’ claim that if we have freewill, it must be true that we could have behaved differently than we did in the past. At first blush, this sounds like a fair description of freewill. After all, if history couldn’t have unfolded in any way other than it did; i.e. everything I did, I *had* to do, it would be a very poor version of freewill indeed, tantamount to saying, “You’re free to have ice cream or cake… as long as you choose cake”, right?

Now, I want to challenge Harris’ assumption about freewill head-on here so let me explicate his claim in a bit more detail. If we have freewill, we could have done other than we did in the past if *exactly the same conditions* hold true. Rewind the tape so that everything in the universe is exactly how it was the first time you made your decision to do X. If you do X every single time we rewind the tape then you were actually compelled to do X and had no choice in the matter. Since we can say exactly the same thing of every action, even those in your future, then you have no choice in anything you do.

This certainly seems like a compelling argument but it is in fact flawed in two ways. First, it relies on the, perhaps intuitive but nevertheless wrong, idea that freewill, if it exists, is something which can only belong to an independent, impartial, stand-alone entity judging each situation in complete isolation from any and everything that preceded it, therefore not predisposed in any one direction. In Harris’ words, “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 those factors.”

This is false. All consciousness is embodied consciousness, engaged in life, limited by a myriad of uncontrollable factors, causing and caused, acting and reacting. Are we completely untainted by “all the factors that determine [our] thoughts and actions”? Of course not. But does the fact that we don’t control *all* of our circumstances mean we don’t control *any* of them? Of course not.

In addition, the very nature of this embodied consciousness is to have preferences, intentions, desires, etc., some, most, or all of which may have arisen in response to external conditions. Noting that this embodied consciousness which I am makes the same decision under *exactly the same conditions* every single time can hardly be a threat to my freedom to make that decision. Wouldn’t it be much stranger if I rewound the tape so that exactly the same conditions prevailed as the first time around (the same intentions, same desire, same state of mind, etc.) and this time I chose differently? To me that would indicate something arbitrary in my ‘free’ choice. In that case, my intentions, desires, will, etc. would seem to count for nothing. That would be a bigger threat to genuine freewill than Harris’ claim.

The second flaw is that we think making the same decision no matter how many times the tape is rewound is equivalent to being *compelled* to make that decision. This doesn’t follow. I kiss my wife every time I leave the house in the morning but this fact, in and of itself, doesn’t mean I am compelled to. There is nothing contradictory in asserting that I am a free being, freely making the same decision to kiss my wife every single time the decision is presented to me.

So ironically, I agree with Harris (again) that the past couldn’t have been other than it was and if we could rewind the tape, everything would unfold again in exactly the same way, but where he expresses surprise that it happens this way *even if* exactly the same conditions hold I maintain that it happens like this *precisely because* exactly the same conditions hold. His opinion that freewill requires that we be able to make different decisions under exactly the same conditions is incoherent. I made the decision I did, not because I am fully determined by a whole host of precedent causes, but because I am a consciousness with concrete and effective desires, intentions, beliefs, etc.

Harris’ second assumption regarding freewill is that we must be the conscious source of our thoughts and actions. He attacks this with the Libet experiments and an appeal to introspection. I have already discussed why the Libet experiments do not do what materialists claim here and summarised Raymond Tallis’ response here and Alfred Mele’s here so I won’t repeat myself regarding this.

Let’s look at Harris’ comments regarding the source of our thoughts. Where do our thoughts come from, he asks? To some extent, he is quite right in claiming that they just ‘pop’ up “from the void”, by which he means without prior conscious input and as the result of purely physical (therefore causal) events in the brain; after all, what could it possibly mean to say we are conscious of our thoughts before we think them?

Actually, I suspect that with this question, Harris is thinking specifically of meditation. If you have ever tried to meditate the first thing you will notice is that (often unwanted and, to some extent, random) thoughts *do* spontaneously appear without any apparent conscious volition at all. Generalising from this situation allows him to conclude that the brain is *always* serving thoughts up and ‘we’ (as far as this word means anything in this context) are reduced to merely witnessing them as they drift by.

Now the first problem here is that even if I grant Harris his thesis that thoughts arise “from the void”, it is 100% illegitimate to generalise from one single abnormal situation to *all* situations. For example, it is not so obvious that thoughts just pop up, undirected and unsought-out, “from the void”, when we are actively engaged in a task and focusing our thoughts on a specific goal or subject. Indeed, the opposite seems to be the case, *I* (whatever that is) very much seem to be controlling and directing my thoughts (despite the fact that in meditation they seem to spring themselves on me).

Rather than leaping to the highly *un*intuitive conclusion that since *some* of our thoughts *sometimes* spring “from the void”, *all* of our thoughts must *always* be unconsciously directed, wouldn’t it make more sense to suppose that when *we* (I promise I will investigate this pronoun shortly) are actively thinking we can control our thoughts but when we ‘let go’ and choose not to corral our thoughts in any particular direction, the architecture of the brain (which is after all, a biological, organic thing, not a machine we can just switch off) continues to tick over in ways that sometimes produce thoughts, much like sleep can produce thoughts (dreams) we don’t actively, consciously desire?

Second, and more importantly, something is very wrong with the entire scenario as Harris has imagined it. When he reflects on his thinking, he imagines that thoughts arise and then he becomes aware of them. But *what* exactly does he imagine has become aware? In other words, who is the *he*? There is no good answer to this without postulating a soul or homunculus in the brain somewhere and that is, in fact, what Harris does consistently throughout his book when he talks of a “conscious witness” or “conscious agent” which a) “I” am, and b) serves as the locus, and de facto definition, of consciousness.

* “The choice was made for me by events in my brain that *I, as the conscious witness* of my thoughts and actions, could not inspect or influence.” [my italics]
* “…your brain has already determined what you will do. *You* then *become conscious* of this “decision” and believe that you are in the process of making it.” [my italics]
* “*I, as the conscious witness of my experience*, no more initiate events in my prefrontal cortex than I cause my heart to beat.” [my italics]
* “You are not in control of your mind – because *you, as a conscious agent*, are only *part* of your mind, living at the mercy of other parts.” [my italics]

So Harris’ entire book is predicated on the following picture of our mental activity: Brain events 🡪 (determine/create) actions/thoughts 🡪 (then) – a “conscious witness” becomes aware of those actions/thoughts before confabulating reasons and generating the illusion of control and freewill.

So what exactly is this “conscious witness” supposed to be? Harris doesn’t elaborate on it but it appears to be the source of our first person subjectivity and our self-reflective awareness. However, in addition to the “conscious witness”, he also claims that “the brain is a physical system, entirely beholden to the laws of nature…” So the glaring inconsistency here is that *Harris’ entire understanding of consciousness (as conscious witness) is predicated on a mysterious “observer”* lurking somewhere in a “*part* of your [supposedly physical] mind” that sounds suspiciously like what Daniel Dennett has called the Cartesian Theatre. Now I know that Harris denies his view implies anything Cartesian but the phrase “I, as the conscious witness of my experience” and its role as the facilitator of conscious experience is what Dennett would, quite correctly in my opinion, call a holdover from the philosophy of Descartes. And what it does in fact do is reduce consciousness to a single, un-extended, causally ineffectual point somehow removed from the physical. I would like to suggest that this whole picture of consciousness; brain events 🡪 thoughts 🡪 conscious witness, is completely wrong and specifically, the “conscious witness”, *in any form*, is pure myth.

Let’s take stock of what we know:

* The physical brain is a necessary precondition for human consciousness as we know it
* In some unknown way this physical, embodied[[2]](#footnote-2) brain generates non-physical phenomena (at the very least, thoughts and the subjective first person experience)
* This non-physical phenomena is somehow capable of reflecting on itself

Now, you might want to object that there are too many ‘unknowns’ and ‘somehows’ in there. Don’t. We need to get comfortable with them because we are right at the frontiers of our understanding when it comes to consciousness. The unknowns aren’t a problem because, although we don’t know *how* a number of things happen, we know *that* they do. If you seriously doubt your thoughts are real or think your first person experience is an illusion, then you’ve effectively ruled yourself out of any rational conversation on consciousness because since you clearly *aren’t* conscious, you can have no idea what it means to *be* conscious. My questions would be better directed at the latest Google chatbot which may, in fact, be more conscious than you.

Back to the story. If we persist in thinking about consciousness in this dualistic fashion of Harris’, it will remain an intractable problem and we will be forced into nonsensical conclusions, such as consciousness is an illusion or thoughts aren’t *about* anything (because how can one clump of matter be about another clump of matter?).

Harris’ dualism manifests in his notion that thoughts are separate from consciousness. Thoughts arise and *then* we (as “conscious witnesses”) become conscious of them. This is false as introspection can uncover. Have you ever been so focused on something that all of your attention was directed into the task at hand? You may have even lost track of time and were certainly not *actively* aware or conscious of your thoughts. Now imagine someone had come up to you and asked what you had just been thinking. Do you think it’s even remotely possible that since you weren’t actively conscious of your thoughts, because your “conscious witness” was completely focused on the external task, you would ever have to reply you didn’t know? Never, right. You are *always* conscious of your thoughts even if you aren’t actively (consciously) focusing on them (while engaged in some activity, for example) or controlling them (as in while daydreaming) or have even forgotten them (as in after having a few too many; beers not thoughts). *It is impossible to have a thought you aren’t, at the same time, conscious of*. To suggest otherwise is to retreat into Cartesian dualism. In order to fully exorcise Descartes’ ghost from our machine, we must, once and for all, dispel the notion that there is any sort of separate “witness” lurking amongst our neurons, pouncing on thoughts as soon as they arise “from the void”. *If we are conscious of our thoughts it is because our thoughts* are *our consciousness*. In other words, we aren’t observing our thoughts, but this is only because we *are* our thoughts.

This means that we need to amend Harris’ picture a little; (embodied) brain events 🡪 consciousness (incl. thoughts). Note the way that our definition of consciousness has also changed here. Harris identified it as a separate “conscious witness” (presumably ‘comprised’ of the subjective experience and (self) aware “I”), but I am now adding thoughts to this package. Consciousness is, then, a tripartite beast:

* The locus of our first person, subjective, experiential take on the world – the ‘what-it-is-like’ aspect of our conscious lives
* Awareness, which includes a reflexive awareness of self
* Thoughts

Now, although I have itemised each ‘aspect’ of consciousness as though it is made up of three things, *none of these aspects exists as a separate and distinct entity from the others*.

So, this gets us closer but still hasn’t escaped a fundamentally dualistic outlook, this time not between thoughts and consciousness, but between consciousness and brain events, or the non-physical and the physical; in other words, Cartesian dualism still haunts our picture of consciousness. Something has still gone awry.

*Detour*

It’s at this point that I would like to take a brief detour to highlight the two options available to us, both of which include significant unknowns (perhaps even ‘unknowables’) but only one of which seems plausible to me.

First, you can start from materialism, the belief that everything in the universe is physical.

Benefit: Consciousness is no longer a problem. Even though we haven’t put all of the pieces together yet, it can’t be anything more exotic than the behaviour and interactions of very tiny pieces of matter.

Drawbacks: It only ‘explains’ consciousness by denying everything that consciousness is. Thoughts, subjectivity, agency, aren’t explained so much as they are ignored or treated as illusory. This means the finding disagrees with the concrete experiments we are conducting in our heads every second of our lives (this isn’t necessarily damning though… unlike my next point).

It also appears incoherent because, while you can deny that we have genuine freewill, it is not so clear that you can deny reality to our thoughts and subjective experiences, *even if these are somehow illusory or not what we think they are*.

And just as importantly, materialism, cannot, *on principle*, explain how these ‘illusions’ that we are experiencing/thinking about things arise in a lump of non-conscious matter in a non-conscious, material universe. Drill as deep as you like into quantum ‘microtubules’ in the brain and fluctuations of quantum this or that, you’ll never find even the briefest flicker of consciousness down there because, according to the materialist theory this option is predicated on, the tiniest bits of matter in the brain can’t be any more conscious than the biggest bits of matter in there either.

‘Faith’ elements: Despite appearing real, consciousness is somehow illusory in some way. In addition, science will one day explain how the lumps of matter in every human head create this persistent illusion of non-physical phenomena.

Second, you can start from consciousness and treat non-physical mental phenomena as real.

Benefit: This is consistent with how our mental lives appear to us.

Drawback: We can’t explain *how* something non-physical arises from something physical.

‘Faith’ element: Even though we are unable to explain how consciousness comes to be, nevertheless, it *is* as it *appears* to be.

Obviously, I consider the latter option the only plausible one of the two. The biggest reason for this is that the idea that our mental phenomena are all illusory in some way seems completely outrageous. Of course, this isn’t a, “I really, really want to believe in freewill so it just *has to* exist” outrageous; rather, it’s a, “Calling thoughts and the feeling I have that ‘it is like something’ to be me illusions just doesn’t make sense. You can’t deny them without admitting their veracity” outrageous.

--- --- ---

Now, with that detour in mind, let’s pick up that thread where I left it with us still being caught in a form of dualism with (physical) brain events somehow ‘generating’ (non-physical) consciousness. If we are absolutely committed to rejecting any sort of Cartesian ‘mind’ or non-physical ‘stuff’ existing independently of physical ‘stuff’ (as we very much should be) then only one option presents itself; (physical) brain events somehow *are* (non-physical) consciousness.

Now, this relation obviously cannot be one of pure identity, of the form A=A. The firing of physical neurons in your brain could not be more different from the image of a non-physical dog in your ‘mind’s eye’ and yet, it is inescapable that they are in fact intimately related. This is where our inquiry must, of necessity, falter – but we always knew this point was coming. We can’t logically justify it or fully understand *how* the conclusion holds but we know *that* it must, because while we can’t decipher consciousness from the ground up, we can faithfully describe it and infer from what is illuminated before us to that which is still shrouded in darkness (as per the reasoning in the detour above).

So, we finally have our revised picture of consciousness in which we can dispense with all ‘arrows’ and simply state; (embodied) brain events/thoughts/subjective experience/(self) awareness. Cartesian dualism, after a protracted battle, has been banished once and for all.

Now, cast your mind way back to the reason we embarked on this voyage in the first place, namely to respond to Harris’ criticism of freewill that our thoughts arise “from the void”. Let me clearly spell out what I take Harris’ argument to be. Since we don’t know (aren’t conscious of) where our thoughts come from, we obviously don’t control them. And if we don’t control our thoughts, they must arise from un-conscious, autonomic processes. If this is true, we, as a person, can’t possibly be free. Our physical brains are just clunking along responding to, and in turn producing, causes all in tune with a grand causal chain that extends all the way back to the Big Bang.

In light of the long discussion I embarked on about the nature of consciousness, I can see two glaring problems with this, centring on the words, “we” and “know”. Let me take “we” first. Harris separates “I” from my thoughts, however we have seen that the “I” which is consciousness, is not a standalone, independent, Cartesian witness. Rather, “I” *includes* our thoughts, our subjective experience and even those pesky brain events. “I” is not some “conscious witness” separate from everything else going on in my brain or thoughts; “I” *am* everything else.

The biggest problem though lies behind the “know”. In arguing that if I don’t *know* (am not conscious of, as Harris would say) the source of my thoughts, I can’t control them and therefore I can’t be free, Harris has mistaken consciousness (to be conscious of something) for knowledge. The nature of knowledge requires that an object be posited for a subject. Since consciousness is self-reflective, this certainly *does* happen in some form. We can make ourselves an object for ourselves (one of the aspects of consciousness we can’t explain but which we equally can’t deny), as in when we reflect on our thoughts or experience of red, but it can’t *always* happen and where it doesn’t happen is precisely where Harris is looking and failing to find it.

This might seem to be a little convenient but if you think about it, it makes perfect sense. Consciousness, once you pare off all the frills, full colour thoughts, richly textured subjective experiences, unimaginably complex networks of neurons firing in exquisite patterns, etc., is ultimately, at its core, a perspective, a vantage point around which the world reveals itself. If that’s true, it is impossible to make this ‘vantage point’ we are, a true, bona fide object for the subject (which we also are); that is, we shouldn’t expect to be able to ‘know’ ourselves at our core.

This is perhaps best seen by way of analogy. The eye cannot see itself precisely because it *is* that which sees. You can look at it in a mirror and see it *as* a thing in the world but you can never see it *as* the thing which sees. Consciousness cannot know itself precisely because it *is* that which knows. You can reflect on (know) consciousness *as* mental events such as thoughts but you can never know it *as* the thing which knows.

So for Harris to say we don’t *know* where our thoughts come from turns out to be a fundamentally confused notion, and therefore a false one, because we *are* those very thoughts. This fact, we can’t *know* the source of our thoughts, is not a problem for us because knowledge is not a pre-requisite of consciousness, in fact, it is the other way around, consciousness is a pre-requisite of knowledge.

Now it’s time to come full circle and hammer that final nail in the causality coffin I started to close earlier. When I left this topic we had established there were some non-physical aspects to ourselves; namely thoughts, self-reflective awareness, and subjective experience, however they were still dependent on the activity of a physical brain, therefore still subject to causality, therefore still determined.

Now though, we have seen that *physical brain events aren’t different from the non-physical events*. It’s all consciousness. We don’t know *how* this is possible, but I have argued this is the best explanation and the only one that doesn’t take us down a road where we have to deny not just freewill but incontrovertible facts about human existence like subjective experience, self-awareness and even thought.

If this is true then the physical brain, *unlike all other matter*, is no longer purely physical but has non-physical components.[[3]](#footnote-3) Two of these non-physical components are particularly relevant here because they effect a break in the physical, causal chain by introducing non-being into the world. These are our self-reflexivity, the ability to turn our conscious minds back on ourselves, and our ability to think.

First, self-reflexivity means that we are able to make of ourselves an object for ourselves. And this means that when I am happy, for example, I’m not *just* happy; rather, I am consciousness *of* being happy. I am able to ‘step back’, as it were, from myself to take a position on myself. We’ve already covered this but what we haven’t looked at is one of the implications that follows. Reflexivity essentially effects a break in being, introducing nothingness, or negation, into the world – I am *not* what I am. I *am* happy but the reflexive nature of consciousness means that I am *not* purely happy. Moreover, this will apply to any state, thought, belief, etc., that you care to name. Whatever I am, whatever describes me at any moment in time, I will always be something *more* than this, precisely because I can ‘step back’ and take a position on it, because I am always consciousness *of* it.

I can only *not* be what I am because consciousness has made it possible for negation, that is, *non-being*, to come into a world otherwise full of being. Everything *is* and *is* positively (all you will ever encounter in the world is being) but suddenly something appears for us that is *not*; something has slipped negation into the world of being. Far from being mere semantics, this becomes crucially important when we realise that causality can only operate on being. Consider the proposition, ‘my happiness causes me to dance’. This will never be true in the same way as the proposition, ‘the white snooker ball causes the red snooker ball to fall into the corner pocket’, is true. The snooker ball example is a true case of causality because there is nothing at play but being, namely, two snooker balls which *are* completely what they *are* and nothing else. Contrast this with my being happy. My happiness can never *cause* me to dance in the same way that the white snooker ball causes the red one to drop into the pocket, because I am *not* totally and completely my happiness; rather, I am *consciousness of* it. Non-being has slipped in between me (as subject) and what I am for myself (as object). Reflexive consciousness is, by definition, outside of any causal chain precisely because it is a negating perspective on the thing that is part of the causal chain; i.e. itself.

Second, our ability to think allows us once more to directly bring non-being into the world. Thought is not bound by the constraints of being. My brain may be confined to my body, but, as the Stoic philosopher Seneca noted, thoughts, on the other hand, are totally free; even time and space themselves cannot contain thought.

This ability to move beyond time and space with our thoughts and investigate worlds which are *not*, in particular, not *yet*, is what I want to focus on here. Every time you think about the future, a reality which is *not* yet, you introduce nothingness, or negation, or non-being, into the world and this again, for precisely the same reasons elucidated above, effects a break with causality.

The fact that I can envisage a (*non*-present) future, once again bringing non-being into a world otherwise full of being, simply cannot be a part of a causal chain in the same way that the white snooker ball *causes* the red one to drop into the corner pocket. We are dealing with two quite distinct aspects of reality. My ability to consider the future is, if one thinks about it, a truly astonishing thing. It is so counter-intuitive that if it didn’t make up a central aspect of our daily lives, as natural to us as breathing, I doubt we would believe such a thing could ever occur. We have noted that the world of being *is*, but this also means that it is always *present*. Tenses have absolutely no meaning in a world of pure being. There is no such thing as past or future in a world where everything *is* and nothing *is* *not*. Consequently, in a world where future (and past) *do* have meaning, that world cannot be pure being, which is to say, it cannot be a purely causal one. Consciousness creates non-being (‘nihilates’) wherever it goes, and in doing so, creates a gap between itself and being, lifting itself above the causal chain.

It’s important to note that I am not arguing that these features of consciousness lead to actions or events that are *un*caused, in the sense of occurring without any preceding conditions – that would be incoherent. Human actions are influenced by past and present events and motivated by future intentions (themselves influenced by past events) but they are not ‘caused’ by them in the same way that snooker balls *cause* others of their kind to drop into corner pockets.

If you have read any of Jean-Paul Sartre’s *Being and Nothingness*, you will recognise more than a passing similarity between his thought and much of what I have said here. The central idea is that the way we typically think about consciousness is completely wrong. It is far more ineffable than science imagines.

I didn’t actually intend to ramble on so much about consciousness but it turned out that it was crucial to forging our path to freewill and rejecting determinism. Determinist, materialist models of the universe, like the one Harris endorses in his book, *Freewill*, are completely inadequate when it comes to explaining consciousness. In his attempt to square the facts of consciousness with a physical universe, Harris ends up creating an epiphenomenal Cartesian “witness” and committing to the idea that freewill is an illusion despite the fact that there is no possibility of ever explaining how such a non-physical illusion could arise in a physical universe.

We don’t know *how* consciousness arises in an otherwise physical universe but we do know *that* it does. Acknowledging this fact and following where it leads carries us into an expanded concept of the “I” that is consciousness and eventually presents us to ourselves as worldly beings, conditioned and influenced by a myriad of factors but also as conscious beings able to step back and take a position on those conditions, influences and even ourselves; beings capable of bringing forth non-being into the world and freeing ourselves from the physical, causal restrictions that bind all other matter. It is in this way that freewill enters into the physical world, not as an a-causal force, but certainly a non-determined one.

1. Harris does say that his argument is not contingent on materialism and it’s true that claiming you are a soul composed of “soul stuff” won’t circumvent Harris’ central claim that you, as the “conscious witness of your experiences”, are fundamentally not responsible for your thoughts. However, since a) I am not making the claim for “soul stuff” and b) the language that Harris repeatedly and consistently uses throughout the book (“laws of nature”, “cause and effect”, “The brain is a physical system, entirely beholden to the laws of nature…”, etc.) makes it abundantly clear that he is a materialist, I see no problem in directing my arguments against materialism. [↑](#footnote-ref-1)
2. This is an important point which I don’t plan to follow up in any detail here. The brain is a necessary precondition for consciousness but we also need a body of some sort. A brain in a vat could never, I suspect, develop consciousness. To have any first person subjectivity at all requires things to interact with; in other words, in the absence of sensory input there is nothing “it is like” to be me. [↑](#footnote-ref-2)
3. Remember this is not a separate non-physical ‘witness’ produced by, and lurking somewhere in, the brain. [↑](#footnote-ref-3)