



# Free Will Project

Martha  
Davey



**SUPERVISED BY HAZEL (OXFORD  
PHILOSOPHY & THEOLOGY GRADUATE AND  
MASTER OF PHILOSOPHY IN  
ANTHROPOLOGY)**

# To what extent can science disprove free will?

## Introduction:

The spoke of a wheel cannot change the course of the cart, however it may envisage alternative routes. In this essay, I will argue that, conceptually speaking, free will can utilise scientific ignorance to escape a definite disproof, but our everyday understanding of freedom can be fundamentally invalidated. I will use a broad and intuitive definition of free which can be easily applied to scientific thinking (based on that of Hossenfelder)<sup>1</sup> – an action is free if it cannot be predicted by all (in principle) available information at the time, nor is it the result of a random or probabilistic process. In practice, this would be shown by a decision which the most highly skilled neuroscientist with a complete map of the person's brain and past could not predict. This decision also could not arise from the random or probabilistic movement of particles in the brain, as although unpredictable, this is not in the person's control. What, then, are we using to try and disprove this notion with – what is 'science'? For the purposes of this essay, a scientific fact is one which can be verified by physical or mathematical experimentation: hence with these definitions we can address the relationship between the two concepts. In this essay I will attempt to show that science can disprove free will to a substantial extent, but never fully.

## Abstract:

The current debate on free will has three main camps: hard determinists (incompatibilists who uphold determinism), libertarians (incompatibilists who uphold free will), and compatibilists (who uphold both determinism and free will). Sam Harris is an example of a hard determinist who draws most of his arguments from neuroscience, as recent developments in this field have supported his view increasingly. For example, Haynes' 2008 experiment, which showed that brain activity occurs hundreds of milliseconds before a conscious decision is made.<sup>2</sup> Many scientists use similar results to show that the person does not make a conscious decision, but their brain makes it for them. Daniel Dennett, a compatibilist, argues that free will remains consistent with these findings due to our ability of self-influence. However the field most threatening to hard determinists is not compatibilism, but quantum physics. Ever since Born explained electron interference patterns with a wave of probability, microscopic physics has taken an indeterministic path: experimental confirmation has further shown that the quantum world is essentially undetermined. This exorcises Laplace's demon (a hypothetical entity within classical physics which knows the exact location and momentum of all particles in the universe and can consequently predict the future precisely). But can this end to classical determinism prove free will? If not, why do we hold the illusion of free will? And can we live without it? All of this will be addressed in detail in the following essay.

## Literary Review:

The first paper I wish to consider is John Conway and Simon Kochen's 'The Free Will Theorem', a physics paper which attempts to show that if an experimenter possesses free will, then fundamental particles must as well. Their conclusion is a possible interpretation of quantum behaviours, but not conclusive proof of free will in my opinion. The paper summarises the theory as follows: (spin related statements are used in the paper, but the statement is a general one)

---

<sup>1</sup> S. Hossenfelder, 'The Free Will Function', preprint: arXiv, 2012 [unpublished]

<sup>2</sup> K. Smith, 'Neuroscience vs Philosophy: Taking aim at free will', *Nature*, vol. 477, 2011

'If the choice of directions in which to perform (spin 1) experiments is not a function of the information accessible to the experimenters, then the responses of the particles are equally not functions of the information accessible to them.'<sup>3</sup>

This partially fulfils my definition of free will; however their hypothesis does not take enough note of probability. It is my view that if a particle acts due to probability, this does not assign it free will, as there is no element of control involved. So, if particles exhibit 'free will' in these spin related experiments, but not in other centrally probabilistic processes, free will cannot be said to be a general property of the particles. Furthermore, what about particle processes occurring without experimental interference, without observation? The same link from person to particle cannot be made in these situations, so if particles have free will it is *very* selectively employed. Therefore I don't believe this argument can be considered a proof of particles' free will, as the precise proof used is a unique case and not a general display of freedom in the universe. Furthermore, as particle physicist Sabine Hossenfelder points out in her rebuttal of this theorem:

'It necessitates free will to make room for free will'<sup>4</sup>

The free will theorem, at most, shows that *if* humans possessed free will it might reflect on the properties of particles too (in some situations). It makes no adequate effort to argue for human free will in the first instance, arguing purely against determinism. This paper thus attempts to prove free will by assuming free will, an approach which I find ineffective in its circular nature.

Remaining in the world of physics - Sabine Hossenfelder's paper 'The Free Will Function' provides a more persuasive attempt to build a capacity for free will into the framework of science. She uses a definition of free will very similar to mine, however crucially it does not reference probability. This leaves her argument vulnerable to quantum mechanical attacks, as it allows scope for quantum events to be included in 'free actions'. Aside from this discrepancy, I find her central thesis to be logical and convincing. The crux of her argument is a 'function' existing outside the realm of differential equations, described as follows:

'One just needs an (in principle) uniquely specified function that cannot be evolved forward for example because it makes discrete jumps or because it is not forward deterministic. That function must in addition have the property that it cannot be constructed just by collecting a (possibly infinite) set of values to the past.'<sup>5</sup>

Her criteria also exclude any randomly generated result, so, overlooking the issue of probability, her function meets the requirements for free will. This proof of a *capacity* for free will relies on the ignorance of current human understanding, she reaches beyond current mathematics to envisage a function which doesn't need a known input to produce an answer. She has effectively chased the definition of free will too far for science to catch. However, her solution has some issues: most importantly, it is an algorithm, not a tangible and human concept. How would this theoretical concept be implemented in the brain, or employed by humans in decision making? Though this abstract notion of free will has proved possible in her logic, this is not the free will people perceive in their lives. Furthermore, she has only proved the *possibility* of a free will function, not found it – as explained earlier, this is a proof of *capacity* to exist, not existence itself. Therefore despite fulfilling the majority of my definition, I am hesitant to call Hossenfelder's function 'free will'.

The area of neuroscience also has great relevance to the question of free will through its recent development. For example, Martin Heisenberg observed that some neurological mechanisms

---

<sup>3</sup> J. Conway, S. Kochen, 'The Free Will Theorem', *Foundations of Physics*, vol. 36, 2006

<sup>4</sup> S. Hossenfelder, 'The Free Will Function', preprint: arXiv, 2012 [unpublished]

<sup>5</sup> Ibid

such as the opening of ion channels in the brain occur entirely at random. This gained the interest of quantum biologists, some of whom believe this unpredictable behaviour might be harnessed to produce free will. On the other hand, thinkers such as Sam Harris perceive neuroscientific results as disproving the idea of free will because they show every action to be an uncontrollable (if not entirely predictable) process. There is evidence that brain activity begins up to four seconds before a conscious choice, which supports this point by suggesting a person merely observes a decision made for them by their brain. Nevertheless, some philosophers find room for free will within cutting-edge neuroscience - one example is Daniel Dennett, who argues for indirect control of our actions through self-improvement. I find Harris' interpretation of neuroscience more convincing as Dennett's reasoning falls prey to the problem of infinite regress, however Harris does present an inaccurate portrayal of the mind's status within the person as a whole.

The core idea behind Dennett's self-improvement argument is one of personal identity, specifically the belief that the current 'character' or identity of a person is the sum of all their past experiences and decisions. Dennett argues that despite a person having no control in the moment of a decision, it is their own past thoughts and choices which dictate their psyche and hence their action. He therefore concludes that we have a 'indirect' free will via our ability to self-educate our own minds, which then make our decisions:

'When one desire triumphs, this is not usually utterly inexplicable, but rather the confirmable result of efforts of self-manipulation and self-education, based on empirical self-exploration.'<sup>6</sup>

I do not find this view entirely persuasive due to the issue of infinite regress, which I believe causes Dennett to commit a logical fallacy. He presents a chain reaction of events which shape and alter a person's character to determine future choices, however this leaves the unanswered question of a first thought. Every child born must have a first thought which cannot have been determined by their past decisions, as they have made none. Consequently this thought must be prescribed by the universal state and its physical laws which positioned all the atoms of the child's brain precisely to construct it. This expands the chain reaction beyond the brain and self-education to a wider list of causes for every decision, going back to the beginning of time and infinitely further still. Therefore, Dennett has simply selected a small number of the infinite causes of an action, which happen to originate from the person's brain. Looking at the bigger picture, self-improvement plays a minor role in determining our decisions, meaning I do not hold Dennett's idea to be a proof of free will.

A possible defence of Dennett's argument would say that even if a first thought is determined by the wider universe, the majority of a person's character comes from their self-influenced past actions and thoughts of which there are many more. I would not view this as a valid defence because I see no reason to give recent causes for an action (such as a past decision) more weight than ancient causes (such as the beginning of the universe), because without either event that action would not have happened. Resultingly, I maintain that self-education plays a very limited role in decision making, thereby not providing scope for free will.

An ulterior interpretation of neuroscience comes from Harris – he asserts that, as every decision can be explained through psychological and physical laws, no one can claim autonomy in their actions:

'Actions, intentions, beliefs, and desires can exist only in a system that is significantly constrained by patterns of behaviour and the laws of stimulus-response.'<sup>7</sup>

---

<sup>6</sup> S. Harris, *Free Will*, New York: Free Press, 2012

<sup>7</sup> D. Dennett, 'Reflections on Sam Harris' "Free Will"', *Rivista Internazionale Di Filosofia E Psicologia*, vol. 8, no. 3, 2017

He finds Dennett's interpretation of these same determining laws invalid because he feels Dennett's concept of free will is not relevant to everyday life. He describes compatibilism as an assignation of ownership over determined actions and no basis for free will:

'Compatibilism amounts to nothing more than an assertion of the following creed: *A puppet is free as long as he loves his strings*'<sup>8</sup>

Harris states that the type of free will he has disproved is the 'popular' one – the idea that if time were rewound to the very same situation a person could act differently. I see this concept of free will as easily overridden by Harris' use of scientific predictability, for if all atoms were in exactly the same place and governed by the same laws, it is logical to predict exactly the same thing will occur each time (bar probabilistic processes). Despite finding this argument logical, I do also perceive some problems within it. For instance, he often compares people to a 'conscious agent' which is simply 'observing' the functioning of the brain. This appears to me to be an incorrect presentation of the mind as our consciousness is a feature of the brain, not some supernatural entity floating above it. Resultingly I disapprove of Harris' isolated portrayal of the mind and prefer Dennett's concept of 'husbandry' or ownership over our thoughts and actions. Aside from this discrepancy though, I judge that Harris provides a valid and logical disproof of 'popular' free will and that Dennett's attempt to redefine free will as a form of indirect control is incorrect.

Thesis:

*Theoretical Free Will -*

Finding Harris' dismissal of 'popular' free will convincing, however respecting the logic of Hossenfelder's free will function, I will argue the first part of my thesis on these premises:

- 1) A free action is one which cannot be predicted from prior states of the universe
- 2) A free action cannot be the result of a random or probabilistic process
- 3) Our current scientific understanding can only explain events based on prior states of the universe, randomness, and probability

Therefore science *cannot* disprove free will, as the criteria for what free will actually *is* can effectively outrun science (as Hossenfelder's free will function does). Free will is inherently *outside science*, and any attempt to fit the two together will inevitably come up against a disproof (as Harris finds). Consequently *if* free will exists it is certainly outside our current understanding, not to say it is extra-physical, but simply requires greater intelligence to comprehend.

Counter arguments to this thesis are found mainly in compatibilist stances which find room for free will inside current scientific knowledge, and I have demonstrated my response to some of these in my analysis of Dennett's views. Essentially though, I find views like Dennett's logically flawed and I believe no allocation of ownership over actions can make them truly free. Therefore a compatibilist definition of free will is, in my view, invalid and cannot effectively challenge my incompatibilist premises. Other forms of compatibilism but Dennett's may prove more effective attacks however, and it would require further research to defend my premises against these.

Further challenges may tackle the third premise, with the development of quantum biology or the discovery of a unifying theory our ability to comprehend a concept outside our current barriers may be possible. If non-differential mathematics is made applicable to a free will function, or string

---

<sup>8</sup> S. Harris, *Free Will*, New York: Free Press, 2012

theory applied to Conway and Kochen's free will theorem, perhaps this abstract free will can be challenged by science. But for now, I believe it is out of science's grasp.

### *'Popular' Free Will -*

To give a rounded analysis of free will I shall also consider the 'popular' concept of free will, which I believe can be fundamentally disproved on the following premises:

- 1) The popular view of free will is one where actions are undetermined, and non-random
- 2) Current levels of scientific understanding show that all actions are controlled by physical laws or random events

This shows that 'popular' free will is incompatible with scientific knowledge due to its demand for direct control by the conscious mind, an idea which can be disproved with modern neuroscience as I have previously described. My reasoning aligns precisely with Harris' arguments but omits the faulty idea of a 'conscious agent' somehow separate from the brain.

Despite this, to fully understand this illusion of 'popular' free will we must address the question of why we have such an illusion. This question would need further research, nevertheless I would like to offer my preliminary answer in two parts (firstly why we have the illusion of choice, and secondly how this gives rise to a belief in free will). To answer the first question:

It is human intelligence that is the cause of our illusion of decision making.

To explain further – humans have a level of understanding which allows us to estimate the reasons behind our actions (such as choosing tea over coffee). Therefore we are also able to imagine a plausible set of reasons for carrying out similar actions (such as choosing coffee over tea). However without the realisation that the precise state of the universe dictates we will choose tea, our reasoning tricks us into believing we 'decided' between two equally possible options. We cannot imagine a set of reasons which would lead us to drink, say, lava in the morning, so we do not perceive ourselves as choosing between lava and tea every day. In fact, the choice between coffee and tea is no more of a decision than that between lava and tea, as tea is the decided outcome of either. It is only our understanding of how we *might* have coffee which creates the illusion of choice.

Now to answer the second question: it is because our habitual operation under such an illusion that humans instinctively believe in free will from early in life. We extrapolate from our supposed experience of 'choice' a wider ability to enact decisions according to our wishes – 'free will'. Disprove the notion of choice, and free will doesn't have a leg to stand on.<sup>9</sup>

### Conclusion:

Areas for further research around these topics would include additional reasoning for why we hold an illusion of free will, and the implications of this on moral responsibility including the use of just punishment in society. In this essay however I have assessed several theoretical and mathematical frameworks used in an attempt to introduce free will into physics, and I have also considered neuroscientific interpretations both in favour of and opposing free will. In my thesis I have shown that theoretical free will is so exceptional in its criteria that it consequently evades scientific capabilities, guarding itself from disproof with our ignorance of non-differential mathematics. In addition, I have placed 'popular' free will firmly within the bounds of science for

---

<sup>9</sup> As free will is an idea derived from another concept I have shown to be illusory (choice); this suggests that it too must be illusory. Imagine a creature with no illusion of choice, seeing the world as a set series of events. Logically this creature would not perceive itself to have a property which allowed it to choose between many alternatives – therefore free will can only arise from an illusion of choice, meaning it itself must be an illusion too.

logical reasoning to invalidate, resulting in the statement that free will as we perceive it is an illusion. To justify this I have rationally assessed the concept of choice to be an illusion and found this fact to necessitate the equally illusionary nature of free will. Considering every form of free will I believe to be valid in this debate (and therefore excluding Dennett's indirect form), I have concluded that our instinctual idea of free will is an illusion of our intelligence, but that the still unknown nature of reality fosters hope of a freedom embedded deep within the fabric of our universe.

## Bibliography:

[Oxford Referencing Style]

Cave. S, 'There's no such thing as free will', *The Atlantic*, 2006, URL:

<https://www.theatlantic.com/magazine/archive/2016/06/theres-no-such-thing-as-free-will/480750/>

Conway. J, Kochen. S, 'The Free Will Theorem', *Foundations of Physics*, vol. 36, p. 1441-1473, 2006

Dennett. D, 'Reflections on Sam Harris' "Free Will"', *Rivista Internazionale Di Filosofia E Psicologia*, vol. 8, no. 3, 2017

Doyle. B, 'The Physics of Free Will': in *Free Will: The scandal in Philosophy*, The Information Philosopher, 2011, p. 223-233

Genovese. M, Myrvold. W, Shimony. A, 'Bell's Theorem', *Stanford Encyclopaedia of Philosophy*, 2021, URL: <https://plato.stanford.edu/entries/bell-theorem/#PhilImpl>

Greene. B, 'Microscopic Weirdness': in *The Elegant Universe*, London: Vintage Books, 1999

Harris. S, *Free Will*, New York: Free Press, 2012

Hossenfelder. S, 'The Free Will Function', preprint: arXiv, 2012 [unpublished]

Kane. R, 'Introduction: The Contours of Contemporary Free Will Debates': in *The Oxford Handbook of Free Will*, 2005

Lemos. J, 'Kane, Pereboom, and Event-Causal Libertarianism', *Philosophia*, vol. 48, no.2, 2020

Muskan. S, 'Hidden Variable interpretation of quantum mechanics', *Quantum Untangled*, 2021, URL: <https://medium.com/quantum-untangled/hidden-variable-interpretation-of-quantum-mechanics-4d1684c3da4a>

Schooler. J, Vohs. K, 'The value of believing in free will', *Psychological Science*, vol. 19, no. 1, 2008

Smith. K, 'Neuroscience vs Philosophy: Taking aim at free will', *Nature*, vol. 477, p. 23-25, 2011