



Article

# De Libero Arbitrio—A Thought-Experiment about the Freedom of Human Will

Johannes Schmidl

Save Energy Austria, Franz Josefs Kai 13/12-13, 1010 Vienna, Austria; johannes.schmidl@chello.at

Received: 8 January 2020; Accepted: 11 February 2020; Published: 16 February 2020



**Abstract:** The discussion of whether or not humans are able to act freely is ongoing, even though, and precisely because, technical methods for detecting the physical state of the brain are constantly improving. The brain as a physical–chemical object seems to be pre-determined by its physical and chemical states, while at the same time human consciousness gives the impression of being able to decide subjectively and freely on its own. Determinists claim that this free decision is just a form of misinterpretation of an epiphenomenon and that the alleged “free decision” has actually been determined by the physical state of the brain before the human subject gives the impression of being able to decide freely. The basis for this is a set of experiments, the first of which was specified by Benjamin Libet. Determinism, as the philosophical position that all events are entirely determined by previously existing causes, in principle enables the existence of a perfect predictor. In this paper, a thought-experiment is introduced which demonstrates that a subjective consciousness can break any forecast about its physical state, independently of the method of its detection, and, consequentially, to refute claims about its purely deterministic role. The thought-experiment picks up on an idea of the philosopher Alvin I. Goldman. Logically, the proof follows the path of a ‘reductio ad absurdum’.

**Keywords:** thought-experiment; libero arbitrio; freedom of will

## 1. Introduction

In an interview for “New Scientist” in February 2011, biological novelist Christian de Duve stated that the cost of human success on earth is the exhaustion of natural resources, leading to energy crises, climate change, pollution and the destruction of our habitat. If we continue in the same direction, humankind is heading for some frightful ordeals, if not extinction. He believes an inherent selfishness in human nature is responsible for this behaviour passed on through human genes. [1]

This somewhat fatalistic statement raises an old philosophical question: are human beings able to act according to free will or are they pre-determined—pre-programmed—by the physical state of their brain, which is based on their genetic traits.

First of all, the set-up, results, and two main conclusions of the famous Libet-experiment [2] have to be described in order to be able to shed light on the context of the thought-experiment:

The set-up of the experiment involves EEGs being attached to subjects’ brains, monitoring when their brains indicate a decision is made and comparing this to when subjects indicate being aware of having made a decision. The results show that the pre-conscious brain indicates a reliably predictive “readiness potential” signal approximately 0.33 s prior to test subjects reporting an awareness of having made a decision on their own. The two main interpretations are:

- First, as the readiness potential reliably predicts the subsequent decision, Libet concluded that the brain has decided prior to having become aware of consciously making a decision. This implies that the conscious awareness of having made a decision is an epiphenomenon, which undermines the free will notion that volition is informed by conscious election.

- Second, Libet was nonetheless of the opinion that this study does not rule out the possibility that individuals retain a “veto” option, that is to say they can override what the brain delivers, if they so choose, and veto the brain’s decision just prior to the moment of enacting it.

More recent work has established, first, that when test subjects were asked to press either a left-hand or right-hand lever, analyses of their brain states reliably predicted which hand would be chosen up to six seconds prior. Secondly, the readiness potential itself has been debunked as a sort of mathematical noise of the equations used to measure brain activity.

This paper links its interpretation of the thought experiment with Libet’s assertion that the conscious agent retains a veto power. We shall assume test subjects would wish to falsify such a prediction every now and then, supporting the existence of free human will.

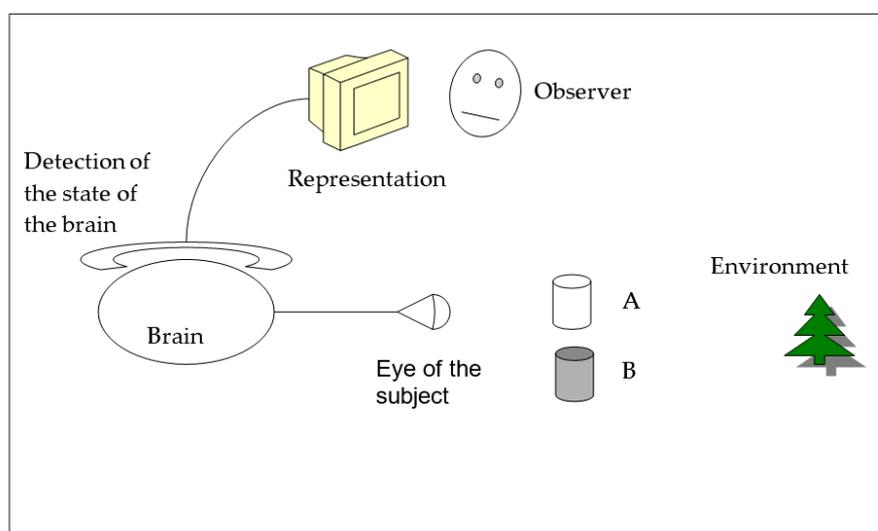
## 2. A Thought-Experiment in Favour of Free Human Will

In preliminary matters, we need one axiom:

A system is considered to be determinate, as long as its state and therefore resulting future can in principle be known in advance. For the following thoughts, it is sufficient if the state and the further dynamic development of the system can be known at least two seconds or more into the future.

We start with the system in Figure 1. The system consists of the following elements:

- a human being, indicated in the figure by its brain and its eye;
- a detecting probe with related software and representation on a screen, which allows for the full detection and representation of the complete physical/chemical/biological state of the brain;
- an observer, usually a human being. The observer can also be placed outside the frame of Figure 1, meaning that it is not decisive that they are part of the system Figure 1, which is considered to be deterministic, or not;
- a simple decision-process for the human being, indicated by the two cups A and B;
- and the environment of the experiment, which can also be part of the deterministic system.



**Figure 1.** A system consisting of a human who has to decide whether to drink a cup of pure water (A) or one of wine (B). Only the brain and eye of the human being is sketched. The brain of the human is fully detected; its state of being is represented on a screen. If the way the human brain works is predetermined, the observer will know in advance if the human will choose possibility A or B.

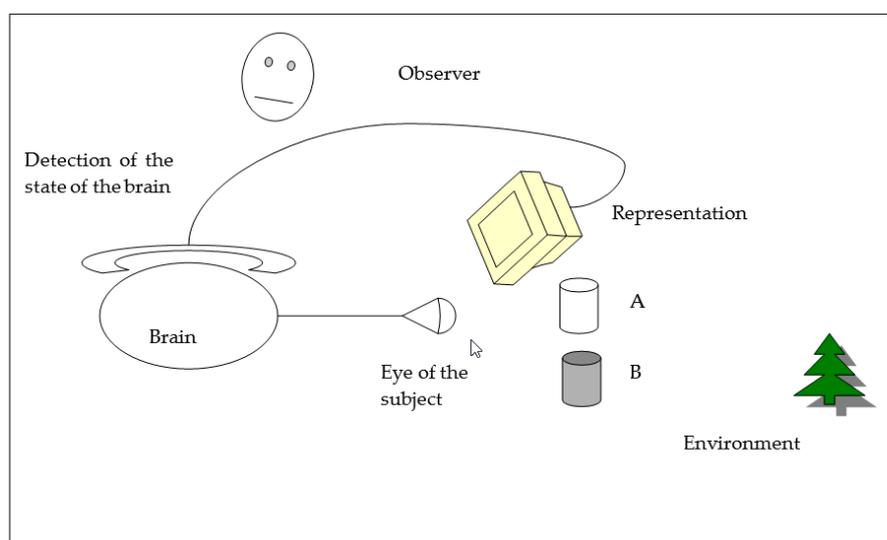
The whole system in Figure 1 is considered to be pre-determined, a classical physical system without in-deterministic influences. The brain of the human being produces electrochemical results which “somehow” are related to subjective perception. Instantaneously, the whole objective system of

the “brain” with all its signals is detected by a probe, interpreted by a computer and is displayed to an observer. The “somehow”-production of subjective impressions as well as the details of the detecting probe and the representation of the state of the brain, are not the focus of this paper.

The observer can, as a consequence of detection, interpretation and representation, know in advance what the electrochemical system “brain” intends to do, when its subjective correlate, the human consciousness, is confronted with a simple decision. The human has to decide between two options with almost the same level of priority, meaning that both are more or less equal for them. In our example, the human has to decide whether to drink from cup A which contains pure water, or from cup B which contains red wine. This is also the principal idea of Libet’s experiment.

The system “brain” is a determined physical/chemical/biological system, as are the rest of all elements in Figure 1. In Figure 1, the brain would make the decision prior to the human’s subjective perception where it actively decides between possibilities A and B. Therefore, the observer would know: “you decide to choose option A” and the human would follow. Everything remains in the predetermined world.

Now, let us change the situation in Figure 1 slightly to that in Figure 2.



**Figure 2.** This is the same system as in Figure 1, except for one difference: now the human actor who has to decide whether to take cup A or B can simultaneously watch a screen which would represent to himself/herself the inner state of his/her brain. The human actor would know at least about two seconds in advance which of the two possibilities, A or B, his/her brain would select. Due to his/her consciousness, however, the human actor is now able to refute any predetermined decision of his/her brain.

We turn the screen representing the inner “physical” state of the human brain around so that the eye of the human being, the brain—which is being detected and observed—is able to watch it (Figure 2). Again, the whole system, as a hypothesis, is considered to be determined.

In this case the human being—again sketched as a brain and eye in Figure 2—would learn about two seconds in advance which decision it is about to choose (A in our example). This again happens prior to the human’s subjective perception that it has actively decided between possibilities A and B. Is the human being now obliged to choose alternative A, as the representation of the inner physical state of his/her brain tells them via the screen that it is about to choose alternative A? Definitely not.

In Figure 2 there is unquestionably something placed in the system which is able to refute any forecast of the deterministic system. Whenever the chain along brain-detection-representation and observation (b-d-r-o) mirrors a certain result, something in between is able to prove that this very result can also be different.

This “something in between” is the free, resistive will (*libero arbitrio*). The hypothesis represented in Figure 2, that the whole system is a pre-determined system, is disproved.

It is important to keep the following in mind:

1. Even if there were to be a logical element between the representation of the inner state of the brain and the brain itself (meaning somewhere “within” the human being), which were able to produce the output non-A from input A (or carry out any other logical operation), this logical element itself would again be part of the pre-determined system within the frame of Figure 2. Therefore, its state and its future development would both be known in advance. The same would occur if the brain itself were to merely produce this logical element. In a classical system, there is no principal hurdle against forecasting the whole state and the whole future development of the system in Figures 1 and 2, deciding whether a human would select alternative A or B.
2. Both systems within the frames of Figures 1 and 2 are classical systems in the way that quantum mechanical effects and related uncertainties do not play a role in them.

### 3. Conclusions

The critical element in Figure 2, which holds the key, is the brain with its resistive will. This brain is both an object and at the same time a related subject. Simultaneously, the subject observes itself as an object. Determinists state that the object “somehow” produces the subject as a correlated epiphenomenal side-effect. The constellation of Figure 2 shows, however, that the subject can disprove the state of its relating object by refuting its predicted future. It is able to disprove decisions that the object seems to have taken.

If anything that is predicted about the system can be refuted, the system turns into a non-predictable one.

In 1970, the philosopher Alvin I. Goldman [3] proposed a thought-experiment of a perfectly informed, accurate deterministic predictive system, predicting that a subject will perform a certain action (say, raise her hand) at time  $t$ . If the subject were aware of that prediction, they can simply refuse to raise their hand at  $t$ . The thought experiment presented here is designed to support the same idea, by having the deterministic predictive system directly monitor brain and environment of a subject and feed the prediction back to them. It places the bases for the deterministic prediction inside the test subject’s own brain and provides the test subject with what counts as neurological biofeedback, together with deterministic knowledge about their own brain, thereby bringing the thought experiment much closer to home for our intuitions about ourselves and our brains in a way that relates more directly to Libet’s veto power claim.

### 4. Further Questions and Remarks

The hypothesis assumes, for the sake of its *reductio ad absurdum* approach, that the system is reliably deterministically predictive. If so, then the system would be able to know whether the test subject wishes to rebut its deterministic predictions and will integrate that information into its predictive processing algorithms. Again, it would mirror these predictions of the inner state of their brain to the human actor, and the human actor would again be able to react with obeying or refuting this prediction. Eventually the potential reaction-time between prediction, presentation and the reaction would become too short for a resistive reaction on the part of the human actor. Therefore, we must define a threshold period for the game of prediction and reaction of, say, half a second or two seconds.

Furthermore, the hypothesis assumes that a free human actor is characterized by the desire to now and then falsify reliable deterministic predictions about their forthcoming behaviour. There is no empirical evidence for that, and it would be hard to find one that avoids the trap of tautology. Rather, this resistive desire is taken as a definition for a free will.

It should be noted that this thought-experiment is akin to the famous Newcombe’s paradox with a perfect predictor. If we were to redesign Newcombe’s paradox in the way that a perfect predictor

would, in addition to our knowledge of the content of the box containing a million dollars or nothing, predict the complete physical/chemical/biological state of the brain of the human player and mirror it back to this very player (as indicated in Figure 2), it would be possible for this player to win the million dollars.

**Funding:** This research received no external funding.

**Acknowledgments:** The author wants to thank colleagues for their valuable time for discussion of the thought-experiment, especially Dieter Bandhauer, Lutz Musner, Andreas Obrecht, and Walter Peissl.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Witchalls, C. Biology Nobelist: Natural Selection Will Destroy Us. Available online: <https://www.newscientist.com/article/mg20928015-400-biology-nobelists-natural-selection-will-destroy-us/> (accessed on 2 January 2019).
2. Libet, B. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.* **1985**, *8*, 529–566. [CrossRef]
3. Goldman, A. *A Theory of Human Action*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1970.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).