A Thought(less) Experiment on Consciousness in Quantum Theory

Itai Agur*

Abstract

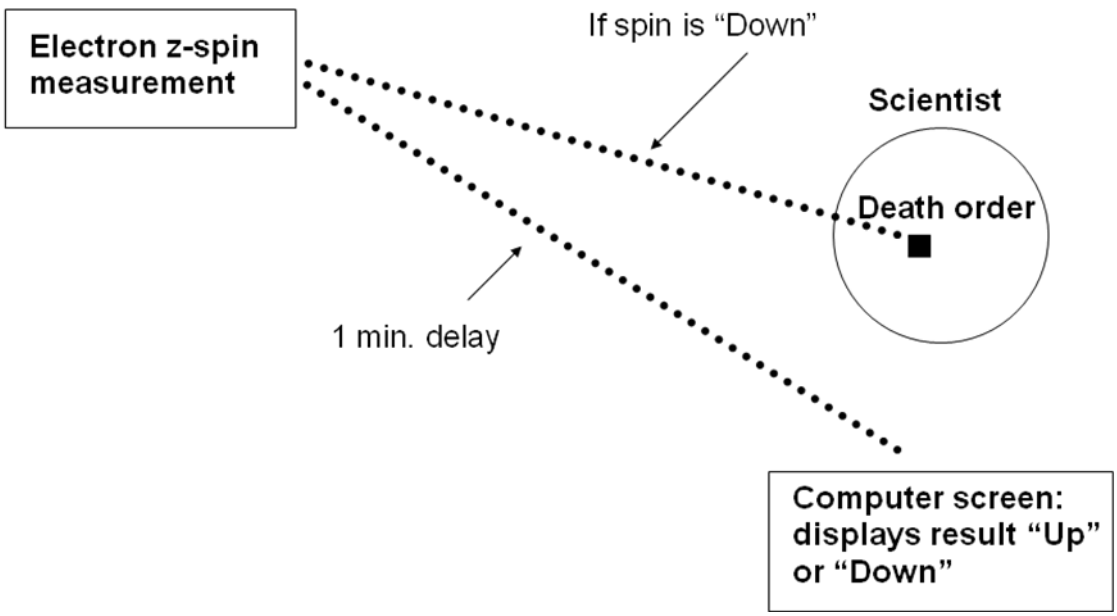
A thought experiment highlights inconsistencies in the view that human consciousness causes wave function collapse. The result of a z-spin measurement is recorded by a computer, which transmits the outcome to a scientist's screen with a time delay. Measuring "Down" makes the computer terminate the scientist's life instantaneously. If the superposition of states ends when the scientist reads the result, then reading "Down" means that he died in the past, before he could have read it. But repeatedly reading only "Up" would be inconsistent with the predictions of quantum theory.

Keywords: Wave function collapse, quantum suicide, observer, interpretation

* itai333@hotmail.com

The view that the collapse of the quantum wave function occurs upon its interaction with human consciousness, finds its origins in the work of Von Neumann [1], and was most famously popularised by Wigner [2,3]. Recently, however, Yu and Nikolić [4] have argued that experimental results from delayed choice quantum eraser experiments (originally proposed by Scully and Drühl [5]) constitute empirical evidence that invalidates the consciousness-based interpretation. This paper provides a thought experiment that adds to this from a different angle, by showing that the consciousness-based interpretation must be either logically impossible or inconsistent with the predictions of quantum theory. The experiment can be seen as a tool to highlight the conceptual difficulties in the consciousness-based interpretation of quantum theory.

Figure 1: Experimental setup



In the experiment, depicted in figure 1, an electron's z-spin measurement is recorded by a computer. The computer has been programmed to send the outcome ("Up" or "Down") to a computer screen, with a one minute delay (added for expositional clarity). Behind the computer screen sits a scientist, next to whom has been placed a hidden bomb, which, if ignited, terminates his life instantaneously.[1] The computer sends a signal to ignite the bomb when it records the result "Down" but not when it records the result "Up". If "Down" is measured then the death signal is sent immediately and thus arrives before the scientist reads the experiment's outcome on his screen. The scientist is unaware of the experimental setup, and as far as he knows he is simply conducting a spin measurement.[2]

The consciousness-based interpretation of quantum theory says that wave function collapse occurs upon interaction of the wave function with human consciousness. Here this means that the superposition of states ("Up" & living / "Down" & dead) collapses to a single state at the moment that the scientist reads his computer screen. If his consciousness collapses the superposition such that "Up" occurred, then there is no problem. But if his consciousness collapses it such that "Down" occurred, then the scientist would have died a minute ago, before reading the screen. Observing the realization "Down" thus erases the timeline in which this realization could have happened.[3] However, if instead the experiment is carried out repeatedly and it always yields "Up", then it is inconsistent with the predictions of quantum theory, which have been abundantly verified empirically: a 50% chance of observing "Up" and a 50% chance of observing "Down".

---

[1] The fact that the bomb shuts down consciousness instantaneously avoids the possibility that the scientist's awareness of dying could trigger wave function collapse. Whether a bomb or other device can really do so is not key, however, since the internal consistency of quantum mechanics cannot be contingent upon man's ability to bring about instantaneous death.

[2] This prevents the scientist from making inferences about the outcome of the spin measurement from awareness of continued living. Although such an inference would not fundamentally alter the experiment, it would make the exposition less clear.

[3] Note that this is different from retro-causality. Retro-causality means that a cause in the present has an effect in the past. Here, instead, a cause in the present brings about the realization of a path that makes the cause in the present impossible.

In other words, if the superposition of states ("Up" & living / "Down" & dead) comes to an end upon the interaction with human consciousness, then time-logic constrains the wave function to collapse in only one way (namely such that "Up" results occur). But such a constraint is inconsistent with quantum theory.

It is important to realize here that the superposition of states can only come to an end at one given moment. Either it collapses when the scientist reads his screen, in which case he cannot be reading "Down", or it collapses at some other point in time, in which case consciousness is not the cause of wave function collapse. Instead, according to some of the prevailing alternative interpretations of quantum theory:

- The wave function collapses upon measurement: the scientist either dies or does not die (Copenhagen interpretation).
- The wave function never collapses, but instead all possibilities materialize: the scientist dies in some universes but not in others (many-worlds interpretation).
- Prior to measurement the electron already had distinct attributes and its z-spin was either "Up" or "Down", not both, and hence the scientist either dies or does not die (hidden variables theory).

The suggested thought experiment is reminiscent of several other well-known thought experiments, namely Schrődinger's Cat, Wigner's Friend and quantum suicide machines [6,7]. Schrődinger's Cat and Wigner's Friend are thought experiments that aim to highlight "absurdities" resulting from the superposition of states. They do not imply logical inconsistencies or observations that contradict the predictions of quantum theory. Marchal [6] applies randomizing destructor-constructor machines in his mechanistic analysis of subjective experience, while Tegmark [7] uses a random-draw suicide machine to highlight the differences between the Copenhagen and the many-worlds interpretations. Instead, this paper applies a similar concept to highlight the potential inconsistencies of the consciousness-based interpretation.

References

[1] J. von Neumann, Mathematical Foundations of Quantum Mechanics, Chap. VI, (1932) translated by R.T. Beyer (Princeton University Press, Princeton, 1955).

[2] M. Esfeld, Stud. Hist. Philos. Sci. B **30**, 145–154 (1999).

[3] E. Wigner, Symmetries and Reflections: Scientific Essays, (Indiana University Press, Bloomington, 1967).

[4] S. Yu and D. Nikolić, Ann. Phys. (Berlin) **523**, 931–938 (2011).

[5] M.O. Scully and K. Drühl, Phys. Rev. A **25**, 2208–2213 (1982).

[6] B. Marchal, Mechanism and Personal Identity, proceedings of WOCFAI 91, M. de Glas and D. Gabbay (eds.), Angkor, Paris (1991).

[7] M. Tegmark, Fortschr. Phys. **46**, 855–862 (1998).