How to Back up or Refute Quantum Theories of Consciousness

Elias Okón and Miguel Ángel Sebastián

Institute for Philosophical Research
National Autonomous University of Mexico
Coyoacán, México

Abstract

Since the early days of quantum theory, both physicists and philosophers have speculated about the idea of a fundamental link between consciousness and quantum mechanics. In particular, it has been suggested that consciousness might be the key to the solution of the quantum measurement problem — the question of deciding under which circumstances, if at all, the wave function collapses. Inspired by this possibility, the thought has been promoted that the quantum level of description is the one at which we should look if we want to provide a satisfactory theory of consciousness.

Is it, however, empirically possible to determine whether or not consciousness is related to the collapse of the wave function? Some have suggested that it is not while others have argued that it is possible to show that they are not related. In this paper we will argue that, even though existing proposals that claim to show that consciousness is not related to collapse do not work (because they are based on a misunderstanding of either the quantum theories in question or the notion of consciousness in play), it is empirically possible to test such claims. Then, based on the fact that any quantum state possesses with certainty a specific property, we will present a concrete empirical way by which the issue could be settled.

1. Introduction

Both consciousness and the foundations of quantum mechanics deeply challenge our understanding of the universe. On the one hand there are, at the very least, prima facie reasons to think that science will not be able to provide a complete explanation of our subjective experience. This has led not few to argue that consciousness falls off the physical order. On the other hand, due to the so-called measurement problem, it seems fair to doubt the coherence of quantum mechanics, the most successful theory we have ever had. Quantum mechanics is incredibly precise in predicting

\[ \text{This is a fully collaborative paper. Authors appear in random order.} \]
the results of empirical measurements but lacks an account of what should count as a “measurement”, a central notion for making sense of the theory itself.

Faced with these issues, many physicists and philosophers have speculated since the birth of quantum theory about the idea of a connection between consciousness and the measurement problem – a good motivation for looking into quantum theory for a fundamental theory of consciousness. According to such consciousness-based interpretations of quantum mechanics, a measurement constitutively depends, some way or other, on the presence of consciousness. Although these views have lost popularity in favor of other proposals – none of them free of unresolved issues – like objective collapse models, Bohmian mechanics or many-worlds scenarios (see Wallace 2008 for a recent review), they continue coming up for discussion. Attempts to settle the debate between consciousness and non-consciousness based interpretations of quantum mechanics have lately been presented in some leading scientific journals. In this paper we show that such attempts fail but we argue that there is, nonetheless, an empirical way to answer the question that divides the two approaches to the measurement problem.

The remaining of the paper is organized as follows: in Sec. 2, we first introduce the well-known problem that consciousness presents for a physicalist understanding of the universe and discuss how, faced with the tension between science and the study of consciousness, some authors have looked into quantum mechanics for an answer. Then, we present the measurement problem and the role consciousness might play in offering a route to a solution. In Sec. 3, we discuss attempts that have been presented in the literature in order to refute the connection between consciousness and the measurement problem and show that they fail due to a misunderstanding of either quantum theory or the notion of consciousness in play. Some authors have gone a step further in claiming that there will always be an irretrievable loss of information in the experiments, and hence, that there is no empirical way to distinguish between consciousness and non-consciousness based answers to the measurement problem.

In Sec. 4, we argue that this latter claim is false and present an empirical way to settle the issue in Sec. 5. We propose an experimental setup that will be able to provide either direct evidence that falsifies the claim that consciousness is necessary for the collapse of the wave function or indirect evidence in favor of the opposite hypothesis. The idea of the experiment is based on the fact that, according to quantum theory, for every system $S$ and property $P$, there is another property $P'$ that has a different value depending on whether $S$ is in a superposition with regard to $P$ or not. We then consider the objection that, due to decoherence, the required measurement is almost impossible to perform in practice and show that very recent satisfactory results in the construction and preservation
of quantum superpositions of distinct macroscopic states suggest that it will be possible to perform such measurement sooner than later. Finally, we conclude by calling attention to the implications of the realization of our proposal for research in consciousness studies.

2. Consciousness, Materialism and Quantum Mechanics

It feels a certain way – or, borrowing Nagel’s expression, there is something it is like – to taste a chocolate cake, to listen Minor Swing or to smell the orange blossom. These are examples of conscious experiences. Conscious experiences are the quintessence of the mind-body problem. Although there is general agreement that conscious experiences – as other mental states – correlate in some way or another with neural activity within the brain, it remains controversial whether and how the grey matter in the brain gives rise to consciousness. Many philosophers accept that there is an irreducible explanatory gap (Levine 1983) between consciousness and matter, between the first-person perspective that consciousness gives us and the third-person perspective offered by our sciences.

Philosophers like Chalmers (1996, 2009) have argued that the right conclusion to be derived from this explanatory gap is an ontological one: conscious experience and physical entities are different in nature. But this opens a new source of problems in explaining the interaction between conscious experiences and the physical world. Alternatively, some philosophers accept the irreducibility of the gap but resist the ontological conclusion, or think that the gap is not irreducible and that future development of our sciences will shed conceptual light on this problem. In this regard, the conceptual revolution that quantum physics has introduced is undoubtedly a suggestive place where to look.

A different way to look at the mind-body problem, with similar results, consists in taking the mind – and consciousness within it – and the physical world as given and wonder about the way in which they interact. A problem arises by the plausible claim that the physical world is causally closed, in which case there is no room for interaction with something outside the physical order. In this framework three possibilities emerge: (i) accept that the mind is causally inert; (ii) accept that the mind is just something physical; or (iii) deny the causal closure of physics.

Accepting (i) requires denying explanations like the one that Mary went for a burger because she was feeling hungry or that we enjoy sex because it is pleasurable. On the other hand, if one accepts (ii) there is no interaction to be explained, but one has to account for the explana-

---

2Chalmers (2003) calls the former “type-B” and the latter “type-C” materialists.
tory gap and derived arguments.\textsuperscript{3} There are reasons to suspect that (iii) rejecting the causal closure of physics is not a satisfactory option. For example, Primas (2002) has rejected the causal closure of physics arguing that the fundamental laws of physics do not determine the initial and boundary conditions required to provide solutions to fundamental equations of motion. It is unclear, however, how this would provide room for the interaction between consciousness and the physical world. Some versions of quantum mechanics seem to open the door for such an interaction, and pioneers of quantum physics like Planck, Bohr, Schrödinger, von Neumann, Pauli, or Wigner considered the role that quantum theory might play for the conflict between physical determinism and conscious free will.

We would like to focus on a different problem essential to quantum theories, which has been one of the main motivations prompting both physicist and philosophers to think of an intimate link between consciousness and quantum mechanics: the measurement problem. This alleged connection has two different sides: on the one hand, some have thought that consciousness might be the key to the solution of the measurement problem; on the other hand, some have speculated that quantum physics might offer new conceptual resources allowing us to formulate new theories of consciousness, and quite often the reason to think so is precisely the measurement problem.

3. Consciousness and the Measurement Problem

The measurement problem, broadly speaking, consists of the fact that, even though standard quantum mechanics depends crucially on the concept of measurement, this concept is nowhere formally defined within the theory. As a consequence, one arrives at a formalism that, in certain circumstances, can become incomplete in an empirically significant way. To see why this is so we start with a few aspects of how quantum mechanics works.

The states of quantum systems are represented by vectors,\textsuperscript{4} denoted by $|A\rangle$, $|B\rangle$, etc., on a type of vector space called a complex Hilbert space (each quantum system gets assigned a specific Hilbert space). In particular, to each possible state of the system corresponds a vector of length

\textsuperscript{3}One can appeal, for example, to the special nature of the concepts we deploy to refer to our experience (Hill and McLaughlin 1999, Loar 1990, Tye 1999 – see Balog 2009 for an excellent review; compare also Chalmers 2007, 2010, Chap. 10).

\textsuperscript{4}We are referring here to so-called pure quantum states, which all closed quantum systems always posses. Quantum systems are sometimes described by so-called mixed states, which are not represented by vectors. Mixed states are used either when the actual pure state of the system is unknown or when the system in question is an entangled subsystem of a larger quantum system.
one, and each vector of length one corresponds to some possible physical state. Now, by definition, the elements of vector spaces can be (i) summed up such that the result is also a vector in the space and (ii) multiplied by numbers such that the result is also a vector in the space. As a result, quantum systems obey the so-called superposition principle, which states that if $|A\rangle$ and $|B\rangle$ are possible states of a quantum system, then any linear combination of them, like $\alpha|A\rangle + \beta|B\rangle$ (with $\alpha$ and $\beta$ two numbers such that $|\alpha|^2 + |\beta|^2 = 1$), is also a possible state of the system.

Such linear combinations are called superpositions. Superpositions are extremely mysterious states, with no classical counterparts, but they are necessary in order to explain observed quantum effects like the interference pattern in double-slit experiments (see Feynman 1994, lecture 6). The important point to stress for now though, which we will explain in detail below, is that a superposition of $|A\rangle$ and $|B\rangle$ such as $\alpha|A\rangle + \beta|B\rangle$ cannot be interpreted, as often has been suggested, by saying that either $|A\rangle$ or $|B\rangle$ is the state of the system – rather, we do not know in which of the two states it will be after measurement (see Albert 1992, Chap. 1).

Next we need to say something about how quantum systems change in time. Standard quantum mechanics contains two radically different time-evolution laws for the state of a system. On the one hand, there is Schrödinger’s evolution, which is continuous, deterministic and linear. On the other hand, there is the reduction or collapse of the quantum state, which is, by contrast, discontinuous, indeterministic and nonlinear. A collapse or reduction, then, is a sudden change from $\alpha|A\rangle + \beta|B\rangle$ into either $|A\rangle$ or $|B\rangle$. In more detail, the reduction postulate states that measurements cause collapses into states of well-defined values for the measured property, with Born’s rule providing the probabilities for the different values obtainable.

Given this state of affairs, a couple of questions arise: How does the theory accommodate this pair of very different evolution laws? Do they give rise to inconsistencies? At first sight it seems that they do not because the standard formalism specifies when to use one or the other law. In particular, it stipulates:

1. When no measurements are taking place, all states evolve according to the Schrödinger equation.

2. When a measurement takes place, states evolve according to the reduction postulate.

This recipe might appear reasonable since it implies that, at each moment, only one of the dynamical laws is at work, thus avoiding inconsistencies. However, a closer look at it reveals deficiencies. The main problem is

---

5 Actually there is more than one such vector; the relation from states to vectors is one-to-many.
that the prescription, which is essential for using quantum mechanics, depends crucially on the notion of measurement, which itself does not have a precise meaning within the formalism. As a result, we obtain, at best, a vague formalism with two incompatible evolution laws, without a clear criterion to decide which of the two must be used at each moment in time. This is, in short, the measurement problem.

In order to respond to this problem, one could point out that, while quantum mechanics deals with very small systems (molecules, atoms, subatomic particles), our measuring apparatuses are very large. Therefore, it seems that there is a way to specify when does the reduction postulate acts, namely, whenever a quantum (microscopic) system interacts with a (macroscopic) measurement apparatus. The proposal then is to claim that measurements are processes that occur only at the macroscopic level.

However, the following question immediately arises: How macroscopic does an object has to be before we can expect its state to collapse? In order to answer this question, let us examine a particular quantum measurement in detail. Following Albert (1992), consider the case of the measurement on a particle, performed by an appropriate measurement apparatus, of a quantum property which we will refer to as “color”; we will assume, as Albert does, that the color of the particle is always measured to be either “black” or “white”.

Now suppose that, initially, the state of the particle is a superposition of the state corresponding to “black” and the state corresponding to “white”. What is going to be the result of the experiment? Well, if we consider the measurement apparatus as such, then we expect the reduction postulate to act so that, at the end, the apparatus will display either “black” or “white”. However, a moment of thought pushes us to acknowledge that the used measurement apparatus, as any other such apparatus for that matter, consists of the same electrons, protons and neutrons described by quantum mechanics. Therefore, we can think of it not as a measurement apparatus but as a quantum object. But if that is the case, the reduction postulate should not act.

As a result, during the experiment, the apparatus should evolve, via Schrödinger’s equation, into a superposition of “black” and “white” – until the display is measured. We could now introduce a new measurement apparatus to measure the display, say a camera, but, of course, we can also treat the camera as a quantum object whose state will collapse until it is measured. It seems then that this argument can be continued in-

---

6We can think of an electron as the particle discussed and the spin along a particular direction as the property to be measured (the “color” of the electron). Spin is a solely quantum-mechanical property with no counterpart (such as angular momentum) in classical mechanics. Since electrons are spin-1/2 particles, spin along a given direction can only have one of two values, “up” or “down”, which would correspond to “black” and “white” in our example.
definitely, without a point at which we can say that a measurement took place. Consequently, there is no criterion to determine when to use the Schrödinger equation and when the reduction postulate.

One can also try to avoid the measurement problem by assuming, along with Bohr, that measuring devices must always be treated classically. However, Bohr’s proposal does not help in solving the measurement problem because it does not provide a well-defined procedure to decide where to draw the line between the quantum and the classical. Besides, it is not clear that the proposal is self-consistent, considering that, as we mentioned above, all measuring devices are made out of quantum constituents. At any rate, almost no one nowadays takes Bohr’s proposal seriously at the fundamental level, so we will assume for the rest of the paper, together with most of the community, that measuring devices must be treated quantum mechanically. This opens up the option that measurement devices may enter superposition states. The standard way to match this with the empirical fact that we never observe such superpositions is through the collapse postulate. But that leads us back to the measurement problem.

A more formal way to present the measurement problem (Maudlin 1995) is by pointing out the mutual incompatibility of the following three statements:

1. The description of the quantum state is complete.
2. Quantum states always evolve according to the Schrödinger equation.
3. Measurements always yield definite results.

This formulation is useful to motivate and classify different possible solutions to the problem. For example, by negating (1) one arrives at so-called hidden variable theories, such as Bohmian mechanics (Bohm 1952), and negating (2) leads to many-worlds scenarios (Everett 1957). In order to negate (3), one needs to specify when the Schrödinger equation is interrupted.

The standard interpretation discussed above lands in this third category of negating (3). However, as we mentioned, it is unsatisfactory because it relies on the undefined notion of measurements in order to state when the Schrödinger equation does not apply. Nevertheless, one can take this third route and do better. One option, advocated by objective collapse models such as GRW (Ghirardi, Rimini, and Weber 1986), is to postulate that collapses happen at random, independently of measurements. Another option is to try to be more precise regarding the notion of measurement. In this regard, a possibility that has been proposed repeatedly throughout the years, is to invoke consciousness for stopping the regress mentioned above (for a review of the different alternatives
proposed in the literature see Okon 2014). That is, we could hold that in order for a measurement to take place, and with it a collapse, some conscious act must be involved.

In this paper we propose an empirical way to test theories that maintain that there is a determinate relation between consciousness and quantum state reduction. For example, Stapp (1993, 1999, 2005, 2006a) has defended that the collapse of the wave function depends on consciousness. On the other hand, Hameroff and Penrose (1996), and more clearly Hameroff and Penrose (2014, p. 67), deny that conscious observation causes quantum state reduction, and rather postulate an identity of the two phenomena. Other quantum theories of consciousness that remain neutral about the relation between consciousness and reduction are not targeted by this paper.⁷

4. The Naive Way of Trying (and Failing) to Refute Quantum Theories of Consciousness

Whether or not consciousness is necessary for quantum reductions to occur seems to be an empirical matter, subject to empirical confirmation or refutation. One might think that in order to test the idea, the empirical setup needed is not that complicated: one could simply seal a quantum system and a measuring device MD in a box, along with a mechanism that allows them to interact only at some given time in the future (say at noon). MD is further equipped with a display that, as soon as a definite result is obtained, shows both the result and the time at which it was obtained. Then, the reasoning goes, if we want to know if consciousness is required for a reduction to occur, we can simply open the box at any time after noon and read the display. If it shows a definite value recorded at noon, then it seems we can conclude that MD was enough to cause the collapse, and so a conscious act is not required for it.

⁷For example, Beck and Eccles (1992), Beck (2001) have argued that quantum processes are involved in exocytosis – the process of releasing transmitters in the presynaptic terminal that initiates the chemical synapsis – without any apparent relation to quantum reduction.

⁸There is, of course, a long history of failed attempts to define what should count as a measuring devise. However, for the purpose of this paper, we do not need to have access to such a definition. All we require from MD is for it to be a “consciousness-free” system (e.g., a system without a brain in Stapp’s proposal, or without living cells containing microtubule-associated proteins in Hameroff and Penrose’s proposal) with different states which are clearly distinguishable by a human observer.

⁹If one is suspicious about the fact that the measurement occurs at a predefined time (because it could introduce a way in which consciousness might be related to the measurement performed by MD), then one can modify the setup mutatis mutandis by introducing a random number generator triggering MD to perform a measurement when a certain sequence obtains.
An experiment in this spirit was proposed by Koch and Hepp (2006), attempting to dismiss quantum theories of consciousness in favor of neurobiological ones. Koch and Hepp utilize the example of binocular rivalry, where one of two eyes is presented with a salient stimulus, like rapidly changing faces, whereas a quantum system is presented to the other. In this paradigm, the subject only sees the salient stimulus, whereas what is presented to the other eye remains invisible and is only rarely consciously seen. For the case in which the considered quantum system is Schrödinger’s cat, Koch and Hepp (2006, p. 612) ask:

What happens to the cat? The conventional prediction would be that as soon as the photons from this quantum system encounter a classical object, such as the retina of the observer, quantum superposition is lost and the cat is either dead or alive.

This is true no matter whether the observer consciously saw the cat in the box or not. If, however, consciousness is truly necessary to resolve the measurement problem, the animal’s fate would remain undecided until that point in time when the cat in the box becomes perceptually dominant to the observer. This seems unlikely but could, at least in principle, be empirically verified.

It seems that Koch and Hepp suggest nothing else than the arrangement we described at the beginning of this section. However, this kind of proposal is based on a naive misunderstanding of the standard interpretation of quantum mechanics, and defenders of quantum theories of consciousness have nothing to fear of them. To illustrate why this is so, consider the Koch-Hepp experiment for the case in which the quantum system to be measured is a particle and MD measures its “color” (see Fig. 1). MD, then, has two displays, one that shows the time at which MD and the particle interact and another one consisting of a needle with three possible positions: Ready, indicating that MD is ready to do a measurement; Black, indicating that the measured particle is black; and White, indicating that it is white. Then, we prepare the particle to be in a superposition of black and white, we arrange things so that the particle goes through MD at noon and we seal the whole thing in a box (A in Fig. 1).

Now, depending on whether or not consciousness is required for measurements to occur, at any time after noon but before the box is opened and examined by an observer, there are two options for the state of the system. The first one (B1 in Fig. 1) corresponds to the possibility that consciousness is not required for a measurement to occur, in which case a collapse happens when the particle goes through MD. Then, after noon, the needle will show the result, say black, and the time display will show the time of measurement. Finally, when a conscious agent observes the system (C1 in Fig. 1), she will find that the time display says 12:00 and that the needle displays Black.
The second option (B2 in Fig. 1) corresponds to the possibility that consciousness is required for measurements to occur, in which case the interaction at noon between the particle and MD does not provoke a collapse. In such a scenario, during their interaction, the particle and MD will evolve according to the Schrödinger equation and the result will be a superposition of the two possible results of the measurement: one in which the needle indicates Black and one in which it indicates White (analogous to Schrödinger’s cat). The important point, however, is that in both terms of the superposition, the time display will indicate the time at which the interaction took place, namely noon. Eventually, then, when the box is opened and observed by a conscious being (C2 in Fig. 1), the state will collapse to only one of those terms and, even though the collapse happened...
much later, the situation for the observer will be indistinguishable from scenario 1.

In reply to Koch and Hepp, Stapp (2006b) makes a similar point but maintains that there is no way to distinguish the moment at which the reduction happens:

According to this conception of quantum theory, the two parallel components of the quantum system will remain superposed until a discriminating conscious experience occurs. This hypothesis is to be contrasted with the common-sense idea that a reduction occurs when the first discriminating macroscopic event occurs. In the words of Heisenberg (1958, p. 54), the transition from “potential” to “actual” “takes place as soon as the interaction of the object with the measuring device, and thereby with the rest of the world, has come into play”. At that point in time all information concerning the quantum phase relationships between the two different parallel components is lost irretrievably into “the rest of the world”, and this implies there is no way to discriminate empirically between the possibility (1) that collapses occur at this earlier point in time, and the possibility (2) that no reduction occurs until some discriminating conscious event occurs.

If Stapp were right, then there would be no way to prove wrong those who maintain that there is an interdependence between the measurement problem and consciousness.

This challenge is faced by Carpenter and Anderson (2006), who acknowledge that there is no way to distinguish between consciousness-based from consciousness-free interpretations of measurement using Schrödinger’s standard thought experiment (described above). However, they claim that it is possible to do so with a more complicated experimental arrangement that codes the quantum outcome of the measurement in two pieces of partial information delivered to two observers. In this way, they claim, it is possible to get “information out of the box, but without an observer being conscious of the quantum state that produced this information” (p. 46).

Making use of the measuring devices introduced in the example above, we can present their idea in more detail. The experiment they propose involves two observers. The first one, $S_1$, sets up the apparatus to give either a true or a false message about the quantum event. That is, $S_1$ decides if the position of the needle is to be correlated or anti-correlated with the true “color” of the particle: i.e. whether, when the “color” of the particle is black, the needle should indicate Black – true information – or White – false information. The second observer, $S_2$, looks at the measuring device and records what the needle indicates. However, since she is unaware of $S_1$’s decision, she cannot infer from what she reads the actual “color” of the particle (that would require information she lacks, i.e., the setup chosen by $S_1$).
The authors claim that this arrangement “allows ... an observer to observe a macroscopic state that is dependent upon a quantum state, as in Schrödinger’s paradigm, but before the quantum state is itself consciously appreciated” (p. 46). Carpenter and Anderson performed an experiment with this setup and observed that neither the state nor the message changed upon $S_1$ becoming conscious of the output of the device. From this they conclude: “our results imply that to collapse a quantum wave-function, measurement alone, rather than conscious observation of a measurement, is sufficient.” The result, then, seems to refute, on one stroke, all consciousness-based interpretations.

Although we agree with Carpenter and Anderson that such theories are subject to empirical refutation, we do not believe their experiment is able to deliver it. The problem with their conclusion, again, is based on a misunderstanding. In particular, they use the expression “being conscious of” as synonym of “knowing that”, and what their experiment shows is that the observers do not need to know the outcome of a quantum detection event in order for a quantum state to collapse. But the actual issue is whether an interaction between consciousness and the device is required. In their experiment there is such an interaction in $S_2$’s observation, even though $S_2$ does not know the state of the system after the measurement. It is at this moment, according to the theories we are considering, when the state collapses.

To see more clearly that the result of Carpenter and Anderson (2006) is not valid, we will show that the predictions of a consciousness-based interpretation are compatible with the actual results of their experiment. From a quantum point of view, when $S_1$ sets up the apparatus (either to give true or false results), the system acquires one out of two possible well-defined quantum states. Such state is known to $S_1$ but unknown to $S_2$. Both of the possible states correspond to superpositions of both possible outcomes of the experiment, but associated in each case with either the right or the wrong message to be delivered to $S_2$ when she measures (when she looks at the needle). At this point, the system, from the perspective of $S_2$, will be modeled by a so-called mixed state that includes two elements of indeterminacy. One of them is due to $S_2$’s ignorance about the setup chosen by $S_1$, and the other one is due to the superposition of the state to be observed (only the second one is related to a quantum effect).

Now, by hypothesis, when $S_2$ makes a measurement, she collapses the state to the term which contains the message she observes. And, importantly, this happens even though she is unaware of its truth or falsity – a well-known quantum phenomenon, present already in the famous thought experiment by Einstein, Podolsky, and Rosen (1935), where local measurements collapse the whole state, even though part of it might be inaccessible to the measurement performed. As a result, one would expect, according to this theory, results identical to those observed in
the experiment of Carpenter and Anderson. In particular, according to consciousness-based interpretations, one would not expect, as they seem to do, the nature of the message to change upon \(S_1\) becoming conscious of the true result.

Now, coming back to Stapp’s reply we want to consider whether decoherence effects – i.e., loss of phase coherence due to the inevitable interaction of any quantum system with its environment – truly cause all the information to be “lost irretrievably” and, hence, whether or not consciousness-free and consciousness-based theories of measurement can be distinguished. In the next section we will argue that they can be distinguished and show an empirical way to discriminate, pace Stapp, scenario 1 from scenario 2. Thereby we will be able to determine whether defenders of the quantum theories of consciousness under consideration make the right kind of predictions.

5. An Empirical Way to Back up or Refute Quantum Theories of Consciousness

In this section we will describe a procedure with which it is possible, at least in principle, to discriminate between a theory that proposes that consciousness is required for collapse and one that holds that collapses happen independently of consciousness. We start by recalling that in quantum mechanics a superposition of, say, states \(\ket{A}\) and \(\ket{B}\) cannot be interpreted as saying that the system is in either state \(\ket{A}\) or state \(\ket{B}\). All we can say is that we do not know in which of the two states it will be after measurement. The reason is that there are measurable properties possessed by the superposition state that are not possessed by either \(\ket{A}\) or \(\ket{B}\) separately.

Therefore, if we need to decide whether the state of a system is in the superposition \(\alpha\ket{A} + \beta\ket{B}\) or in either \(\ket{A}\) or \(\ket{B}\), we can measure the system to see if such measurable properties of the superposition obtain or not. Moreover, we know that any quantum state possesses with certainty a specific property,\(^\text{10}\) so if the state of a system is known, there is a property for which, if it is measured, we are sure to obtain a particular value that can be predicted with certainty. In order to distinguish \(\alpha\ket{A} + \beta\ket{B}\) from either \(\ket{A}\) or \(\ket{B}\), we can measure such a property, and if the state is the superposition state we will necessarily obtain the corresponding

\(^{10}\)This is so because any quantum state is an eigenstate of some observable, from which it follows that the probability of finding, as a result of measuring such observable, the corresponding eigenvalue is 1. In fact, given some state \(\ket{\psi}\), there are many observables for which such state is an eigenstate. A particularly simple observable that does the job is the projector \(P(\psi) = \ket{\psi}\bra{\psi}\) which has \(\ket{\psi}\) as an eigenstate with eigenvalue 1 and any state orthogonal to \(\ket{\psi}\) as an eigenstate with eigenvalue 0.
predictable value. If, on the other hand, the state is either $|A\rangle$ or $|B\rangle$, we will obtain different results.\footnote{Actually, even if the state is either $|A\rangle$ or $|B\rangle$, there is a probability smaller than one to obtain as a result of the measurement the value that one would get if the state were the superposition. Therefore, in order to discriminate between the two scenarios one needs to perform the measurement on a number of identically prepared systems. As soon as one obtains a value different from the one associated with the superposition, one can claim that the state was either $|A\rangle$ or $|B\rangle$ and not the superposition.}

To see how all this works in detail, let us start by applying the procedure to a particle (below we will apply it to the whole box containing the particle and MD). Imagine that we want to know whether the particle is in a superposition of black and white. Measuring the “color” would not work because we know that if we measure the “color” of its state it will automatically collapse into either black or white. However, as we have just seen, there must be another property for which, if we measure it, we will know with certainty whether the particle was in a superposition or not. Following Albert (1992) once more, we will call this property “hardness”, and its two possible results “hard” and “soft”.\footnote{If our particle were an electron and “color” is analogous to the spin along direction $z$, we can think of “hardness” as the spin along some direction different from $z$, and of the value “hard” as spin up along that direction. In fact, all states of a spin-$1/2$ particle can be written as a linear combination of “spin up along $z$” and “spin down along $z$” (i.e., such vectors form a basis of the corresponding Hilbert space). Moreover, all such states have the definite value of spin up for spin measured along some direction. Therefore, for any state there is a direction such that, if the spin along this direction is measured, the result will be spin up with certainty. For instance, the linear combination of “spin up along $z$” and “spin down along $z$” with both coefficients equal to $1/\sqrt{2}$ is the state “spin up along $z$”.

We will further assume that the particular superposition we are dealing with corresponds to a state with a well-defined value of “hardness” corresponding to hard. Therefore, if we want to know if the particle is in a superposition, instead of measuring its “color”, we can measure its “hardness”, and if the result is not hard we can conclude that the particle was not in a superposition. Applying this to the pair particle plus MD, instead of only the particle, is the procedure we are proposing in order to determine whether consciousness is required for collapse. The details of the proposal are given next.

Consider again a single particle whose state is known to be a superposition of black and white, enclosed in a box with a MD. Remember that MD is taken to be a “consciousness-free” system with states corresponding to different outcomes which are macroscopically distinct, i.e., clearly distinguishable by a human observer. As before, MD and the particle are arranged to interact at noon (see Fig. 2). If such interaction does not provoke a collapse – because consciousness is required for that – the state will evolve according to the Schrödinger equation into a superposition yielding different measurement results (B2 in Fig. 2): Upon one result the particle
is black and MD displays this, upon the other result the particle is white and MD displays this. Call this state $|A\oplus B\rangle$.

If, on the other hand, the interaction does provoke a collapse – no consciousness required – then the state of the system will either be one in which the “color” of the particle is black and MD displays this, or one in which the “color” of the particle is white and MD displays this (B1 in Fig. 2). Call these states $|A\rangle$ and $|B\rangle$, respectively. At this point, we can consider MD together with the particle as a single system, and call it The System (TS). So $|A\rangle$, $|B\rangle$ and $|A\oplus B\rangle$ are the possible states of TS. Those who defend that consciousness is required for a collapse maintain that TS is in state $|A\oplus B\rangle$, whereas those who deny that would maintain that it is either in $|A\rangle$ or in $|B\rangle$. Therefore, if we could find out whether TS is
in a superposition or not we could settle the issue. In order to do so, we apply to TS the same procedure we applied above to the particle.

The next step, then, is to bring in a conscious observer to measure TS and, in particular, to measure the specific property of TS that characterizes the superposition state \(|A \oplus B\rangle\) (i.e., the property analogous to “hardness” in the particle example). Clearly, such property is not simply the position of the needle since that would be analogous to measuring the “color” of the particle, which we saw was of no use in order to decide if its state was a superposition. Hence, a more complicated property of TS will have to be measured; we will call this property “temperature” and denote by “\(70^\circ\)” the value of temperature that characterizes the superposition, (i.e., the value we would expect to find with certainty if TS were in state \(|A \oplus B\rangle\)). Now, all we need to do is to ask the conscious observer to measure the “temperature” of TS (C in Fig. 2), and the result of this measurement will reveal whether the state of TS is still a superposition or not.

If TS is no longer in a superposition (the value of “temperature” is not found to be \(70^\circ\)), then a collapse did occur when MD and the system interacted, meaning that consciousness is not required for the collapse to occur. If, on the other hand, the value of “temperature” is always found to be \(70^\circ\), then a collapse did not occur when MD interacted with the system. In this case, a theory as proposed by Hameroff and Penrose (1996, 2014) would be falsified because it predicts that the reduction of the state of TS should happen long before the conscious observation occurs.\(^{14}\) This is, of course, insufficient for showing that consciousness is required for the collapse. However, if we repeat the experiment with very different MDs and the results are that the interaction with none of them gives rise to a collapse, then we would have good reasons to think that something inherently human brings about the collapse – we do not observe superpositions. Consciousness seems to be an excellent candidate in this regard. Summing up, the proposed experiment would either provide

\(^{13}\)We remind the reader that our objective is to propose an experiment that discriminates between two types of collapse theories, those that involve consciousness and those that do not.

\(^{14}\)In their reply to Koch and Hepp (2006), Hameroff and Penrose (2014, p. 67), our italics) make this commitment explicit: “Koch and Hepp challenged Orch OR with a thought experiment, describing a person observing a superposition of a cat both dead and alive with one eye, the other eye distracted by a series of images (‘binocular rivalry’). Without explaining how an observable superposition of this kind could be prepared (where according to OR, by \(\tau \approx h/E_G\) the cat would already be either dead or alive long before being observed), they asked ‘Where in the observer’s brain would reduction occur?’, apparently assuming Orch OR followed the version of the Copenhagen interpretation in which conscious observation, in effect, causes quantum state reduction (placing consciousness outside science). This is precisely the opposite of Orch OR in which consciousness is the orchestrated quantum state reduction given by OR.”
direct evidence against the claim that consciousness is necessary for the collapse of the wave function or indirect evidence that it is not.

One might object that, due to decoherence, the required measurement is almost impossible to perform in practice. This is because TS, being macroscopic, interacts strongly with its environment (for example, with the multitude of particles surrounding it); and as soon as such an interaction occurs, the temperature measurement we propose stops being a reliable method to determine whether or not the state of TS is $|A \oplus B\rangle$.

This is due to the fact that, after an interaction with even a single particle in the environment, it might no longer be possible to assign a pure quantum state to TS (just as it is impossible to assign a pure quantum state to the original particle after its interaction with MD). In this case, a measurement of “temperature” of 70° will no longer indicate that TS is in state $|A \oplus B\rangle$. Thus, we need to make sure that a state like $|A \oplus B\rangle$ is maintained for a period of time sufficiently long to measure the “temperature”.

Difficult as this may seem, amazing advances in the construction and preservation of quantum superpositions of distinct macroscopic states, such as the one of our proposal, have been achieved lately. For example, Friedman et al. (2000) present experimental evidence that a superconducting quantum interference device can be maintained in a superposition of two macroscopically distinct magnetic-flux states. Moreover, Bruno et al. (2013) and Lvovsky et al. (2013) construct a superposition of two macroscopically distinct states of over a hundred million photons – a clearly visible macroscopic entity – resulting from their interaction with a single photon. Note that this type of interaction is precisely what our experiment requires: the stream of millions of photons could play the role of the needle in MD in a measurement of a property, the “color”, of the single photon. That is, millions of photons would play the role of the macroscopic MD, the single photon the role of the microscopic quantum system, and together they would form TS which needs to be measured in order to assess if indeed a conscious observer is required for the collapse.

One might object that the leap from the above mentioned experiments involving photons to ordinary macroscopic measuring devices is vast, and that of course is true. Note, however, that in order for our proposal to work we need it to work for just one macroscopic MD. One might also object that, given that human brains are warm and wet, one must assign

---

15At this point it is important to stress that the widespread belief that decoherence by itself is enough to solve the measurement problem is in fact false. The most common mistake in this regard arises from assigning an incorrect physical meaning to the reduced density matrix of a quantum subsystem. In particular, even if the reduced density matrix in question has the form of an improper mixture, it does not follow that the physical situation of the subsystem is identical with that of an ensemble described by the same improper mixture.
to them mixed states rather than pure states. Note, however, that in order for our proposal to work one needs to assign a pure state to TS, not to the brain that measures it at the end. We conclude from all this that, in the foreseeable future, the proposed experiment could allow us to either refute or confirm consciousness-based interpretations of quantum mechanics, not only in principle but even in practice.

6. Conclusion

The idea that consciousness plays a key role to determine when measurements occur, and hence to control when collapses of quantum states happen, has been repeatedly offered throughout the years. As a consequence, several attempts to dismiss this hypothesis have been presented. We have shown that these attempts typically fail because they are based on a misunderstanding of the theoretical postulates involved. In this paper we have proposed a novel, empirically sound way to determine whether consciousness is involved in measurements or whether collapses can happen independently of consciousness.

If, in our proposal, the interaction of a system in a superposition state with a measurement device results in a determinate state, one can conclude that consciousness is not required for collapses to occur. If so, one important motivation for looking into quantum mechanics for a theory of consciousness would be ruled out, and some particular theories would be immediately falsified. If, on the other hand, one were to always find, upon measurement, the correct value for the property that the superposition state possesses with certainty, then this would give us good reasons to think that consciousness is required for quantum measurements. It would urge us to look further into the quantum realm in order to construct a theory of consciousness.

Acknowledgments

We would like to thank Manolo Martínez and an anonymous referee for their comments on a previous draft. Financial support for this project was provided by DGAPA-UNAM projects IA400114 and IA400615.

References


Hill C.S. and Mclaughlin B.P. (1999): There are fewer things in reality than are
dreamt of in Chalmers’s philosophy. Philosophy and Phenomenological Research
59(2), 445–454.
611–612.
sophical Quarterly 64, 354–361.
Mexicana de Fisica E 60, 130–140.
Primas H. (2002): Hidden determinism, probability, and time’s arrow. In Be-
tween Chance and Choice, ed. by H. Atmanspacher and R.C. Bishop, Imprint
of Consciousness Studies 6(8-9), 143–164.
Journal of Consciousness Studies 12(11), 43–58.
www-physics.
lbl.gov/~stapp/koch-hepp.doc
Tye M. (1999): Phenomenal consciousness: The explanatory gap as a cognitive
In The Ashgate Companion to Contemporary Philosophy of Physics, ed. by
D. Rickles, Ashgate, Oxon, pp. 16–88.

Received: 28 September 2015
Revised: 11 November 2015
Accepted: 08 December 2015
Reviewed by Thomas Filk and another, anonymous, referee.
When I received the preceding article by Okón and Sebastián for review I was both curious and skeptical with respect to the proposed ideas. After reading the article I realized my dilemma more clearly: The text was very well written, the argumentation was absolutely transparent, and I could agree with almost each and every statement by the authors – except their final and main conclusion. This conclusion is that, before a conscious being has read out the result of a measurement, we might be able to measure a superposition of a measuring device and a quantum system, so that obviously the measuring device by itself is not responsible for the collapse of the quantum state. I suggested major revisions with respect to this conclusion which, however, were only partially taken over by the authors, and in some respect even in contrast to what I had in mind.

As I still considered the article to be an excellent contribution to the general discussion about the role of consciousness in certain quantum theories, but on the other hand could not agree with the main conclusion of the authors, I suggested to the editor of Mind and Matter the option now realized: to publish the article together with a commentary in which I try to lay out my position in this matter, thereby disclosing my identity as a reviewer. In addition, the authors have been given the option to respond to my reply at the end of their paper.

My commentary is a summary of my two referee reports in reaction to their original and revised submission. It is my hope that readers, particularly non-experts in this field, who are interested in an impression about the various arguments will profit from comparing two opposing conclusions which are based on very similar opinions about essential aspects of quantum theory.

In my first report I wrote that the authors are correct about the following: If identical measurements are performed on an ensemble of identically prepared systems, and these measurements essentially yield two results $a$ and $b$ (sometimes $a$ and sometimes $b$), we cannot tell whether the ensemble represents a mixture of systems, some of which are in a pure state $|A\rangle$ and some in a pure state $|B\rangle$ (where $|A\rangle$ gives rise to result $a$ and $|B\rangle$ to result $b$), or whether this ensemble represents a pure state $|A \oplus B\rangle$, symbolizing a superposition of $|A\rangle$ and $|B\rangle$. In order to distinguish between these two cases, it is required that we can perform a complementary measurement for which $|A \oplus B\rangle$ is an eigenstate with eigenvalue $c$. For such measurements we would still obtain results $a'$ and $b'$ (sometimes $a'$ and
sometimes \( b' \), which now may differ from \( a \) and \( b \) in the first case above. However, we would always obtain the same result \( c \) in the second case above. We have to perform measurements of complementary observables to distinguish between a mixture (the first case) and a pure state that is a superposition (the second case) with respect to eigenstates of the first observable.

The authors suggest to make use of this distinction for a discussion of the role of consciousness in the measurement process in the following way: Suppose a quantum system QS and a device MD (which can be read as “measuring device” or as “macroscopic device”, see below) have interacted such that the state of MD has changed in correlation with QS. This situation would lead to an entangled state of the total system, “QS being in state \(|A\rangle\) and MD in state \(|a\rangle\)” plus “QS being in state \(|B\rangle\) and MD in state \(|b\rangle\)”.

Now follows the crucial argument of the authors: If consciousness is responsible for the collapse, we should be able to find a different type of measurement (performed before any conscious being read out the result of MD) for which this entangled state is an eigenstate and which, therefore, always yields the same result. If, on the other hand, the macroscopic nature of MD is responsible for the collapse rather than consciousness, such an observable cannot be found.

My objection against their conclusion is this: If we can show that QS plus MD are in a superposition state \(|A \oplus B\rangle\) and not in a mixture of \(|A\rangle\) and \(|B\rangle\) (where \(|A\rangle\) and \(|B\rangle\) essentially correspond to the pointer basis of MD), then this simply proves that MD cannot be regarded as a measuring device. Therefore, it does not follow that the collapse is driven by consciousness. In this same spirit, Bohr emphasized on numerous occasions that we must be able to talk about the results of measurements in the language of classical physics. Therefore (according to Bohr and other proponents of the standard interpretation of quantum theory), a measuring device is defined by the fact that it cannot be in a superposition state with respect to its pointer basis.

In this respect I disagree with the assessment of the authors that “almost no one nowadays takes Bohr’s proposal seriously at the fundamental level”. For Bohr there existed no measuring device at a level where only a few particles or simple quantum states are involved. The authors’ remark that “measuring devices must be treated quantum mechanically” is almost a contradiction in itself: Macroscopic devices may be treated quantum mechanically, but if quantum effects are relevant (e.g., because these devices are observed to be in superposition states with respect to localized macroscopic pointers), they should not be called measuring devices. I agree with John Bell that the term “measurement” should not be used at a fundamental level.

As long as we do not have a better definition of a measuring device, we cannot apply the criterion of the authors. Other approaches that explain
the collapse of a quantum state by an external influence (such as Ghirardi et al. 1985, Károlyházy 1966, or Penrose 1998) specify bounds for the size of systems beyond which their proposal differs from the predictions of the standard interpretation. These bounds basically define what should be regarded as “macroscopic” and what as a “measuring” device, and can in fact be used for tests of the criterion described above.

The key to the issue is that a “measuring device” is not the same as a “macroscopic device”, at least if “macroscopic” refers to the number of particles involved. Gottfried and Yan (2003) define an experimental arrangement as a measuring device if and only if the different pointer basis states are macroscopically distinguishable. They explicitly mention superfluid helium as a macroscopic system which does not function as a measuring device.

So, in the end, the argument of the authors seems to prove exactly the opposite of what they claim in their article: If QS plus MD are found in a superposition state $|A \oplus B\rangle$, this cannot be regarded as a measurement, and if they are always found in a mixture of $|A\rangle$ and $|B\rangle$, as required by a measuring device, this (according to the author) would prove that consciousness is not the reason for the collapse. My conclusion is that, based on the criterion of the authors, almost by definition consciousness cannot be the reason for a collapse.

In their revised manuscript, the authors implemented some changes leading to a new claim: “We present an experimental setup that will be able to provide either direct evidence that falsifies the claim that consciousness is necessary for the collapse of the wave function or indirect evidence in favor of the opposite hypothesis.” My reply to these changes amounted to almost exactly the opposite: If QS plus MD (quantum system plus macroscopic device) is never found to be in a superposition of “classical” states (i.e., well-defined pointer states), this only proves that a corresponding experiment does not resolve the superposition states. This is the experience physicists have made in all measurement processes investigated so far. So, if the authors were right, the proof has already been given.

I think the situation is rather the other way around: If we find a superposition state for QS plus MD (quantum system plus macroscopic device), and MD involves a system of the size of a human being (but without consciousness) with comparatively many degrees of freedom at room temperature, then this is strong evidence that consciousness may be required for the collapse. But, as mentioned in my first report, one may still argue that MD is not a measuring device as required by standard quantum mechanics.

The only way to resolve the issue is to actually include a conscious being: If MD consists of a conscious being and we find the total system QS plus MD in a superposition state (of course, before we ask the conscious
being about the observed result), then this proves that consciousness is not the reason for the collapse. Presumably, such an experiment is impossible. My conclusion: unless we are able to perform quantum experiments involving conscious beings (in the way mentioned above), the issue cannot be settled by proposals of the kind presented by Okón and Sebastián.

References
We are very grateful to Filk for his review and reply to our paper. We believe that the disagreement arises due to a difference in the use of the term “measuring device”. While for Filk, by definition, such an object cannot be in a superposition of pointer states, for us it is merely a “consciousness-free” device whose pointer states are macroscopically distinguishable (such as a Stern-Gerlach device). Let us call the former a measuring device and the latter an MD. Then, at least in principle, we allow MDs to enter a superposition. However, it is an empirical fact that MDs are never observed to be in such states: when a silver atom passes through a Stern-Gerlach device, we observe a spot either up or down on the screen, and never a different result.

There are two possible explanations relevant in the framework of the discussion: either MDs are never found to be in a superposition because when they interact with a microscopic quantum system (a silver atom in the example) the wave function collapses; or although they can be in a superposition, when we observe them, our consciousness (or at least something beyond the MD) makes the wave function collapse. The key question, then, is if MDs are, all by themselves, measuring devices, or if measuring devices require the presence of consciousness.

The experiment we propose is able to settle the question. If the system composed by MD and the quantum system is not found in a superposition, then the MDs are measuring devices and consciousness is not required for the wave function to collapse. If, on the other hand, the system composed by MD and the quantum system is found in a superposition, then consciousness would seem to be involved in the collapse of the wave function.