How to (Back Up or) Refute (certain) Quantum Theories of Consciousness

Elias Okón and Miguel Ángel Sebastián[*]

Abstract

Since the early days of quantum theory, both physicists and philosophers have speculated about the idea of there being 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 that the quantum level of description is the one at which we should look for if we want to provide a satisfactory theory of consciousness has been promoted.

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 existent 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.

Both consciousness and the foundations of quantum mechanics deeply

[*]This is a fully collaborative paper, authors appear in alphabetical order.

challenge our physical 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 seem fair to doubt about the coherence of quantum mechanics, the most successful theory we have ever had. Quantum mechanics is incredibly precise in predicting 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 there being a connection between consciousness and the measurement problem—a good motivation for looking into quantum theory for a fundamental theory of consciousness. According to these 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 been lately presented in some leading scientific jounals. 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 these two approaches to the measurement problem.

The remaining of the paper is organized as follows: in section 1, 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 section 2, 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 misundertanding 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 irretrivable loss of information in the experiments, and hence, that there is no empirical way to distinguish between consciousness and non-cosciousness based answers to the measurement problem. In section 3, we argue that such a claim is false and present an empirical way to settle the issue. We present an experimental set up that will be able to provide either direct evidence that falsifies the claim that consciousnes 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', which 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.

1 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 a 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.[1] In this regard, the conceptual revolution that quantum physics has introduced is undoubtedly a suggestive place in which to search.

A different way to look at the mind-body problem, with similar results, consist 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 the truth of 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 explanatory gap and derived arguments.[2] There are reasons to suspect that (iii) is not a satisfac-

[1] Chalmers [2003] calls the former 'type-B' materialists and the latter' type-C'.

[2] 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]

tory alternative. 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 make 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 Newman, Pauli or Wigner considered the role that quantum theory might play in reconsidering 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 that has led 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 for the measurement problem; on the other hand, some have speculated that quantum physics might offer new conceptual resources from which we could formulate new theories of consciousness, and quite often the reason to think so is precisely the measurement problem.

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, such notion is never 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 by saying a few things about how quantum mechanics works.

Possible states of quantum systems are represented by vectors,[3] denoted

for an excellent review. Cf. Chalmers [2007, 2010, ch. 10]).

[3] We are referring here to so-called pure quantum states, which all closed quantum

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 one,[4] and each vector of length one corresponds to some possible physical state. Now, vector spaces, by definition, are such that their elements can be i) summed such that the result is also a vector of the space and ii) multiplied by numbers such that the result is also a vector of 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 α and β two numbers such that $|\alpha|^2 + |\beta|^2 = 1$), is also a possible state of the system. Such linear combination 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, and which we will explain in detail below, is the fact 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, as saying that either $|A\rangle$ or $|B\rangle$ is the state of the system but that we do not know which is the case (see Albert 1992, chapter 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 postulate,* which is, in contrast, discontinuous, indeterministic and non-linear. A collapse or reduction of the quantum state, then, is a sudden change from, for example, $\alpha|A\rangle + \beta|B\rangle$, into either

systems always posses. Quantum systems are sometimes described with 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.

[4] Actually more then one; the relation from states to vectors is one to many.

$|A\rangle$ or $|B\rangle$. In more detail, the postulate holds that *measurements* cause collapses into states of well defined values for the measured property (with the so-called Born's rule providing the probabilities for different values to obtain). Given this state of affairs, a couple of question arise: how does the theory accommodate this pair of very different evolution laws?, do not 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. 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, looking closer at it reveals its deficiencies. The problem is that the prescription, which is essential in order to use quantum mechanics, depends crucially on the notion of measurement, but such notion 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 of time. This is, in short, the measurement problem.

In order to try to respond to the argument given above, one could point out that, while quantum mechanics deals with very small systems (molecules, atoms, sub-atomic particles), our measuring apparatuses are, in contrast, enormous. Therefore, it seems, after all, 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 try to answer, let's examine in some detail a particular quantum measurement. For example, 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".[5] 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, is built out 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 displaying Black and White—that, of course, until the display is measured. We could now introduce a new measurement apparatus to measure the display, let's say a camera, but, of course, we can also treat the camera as a quantum object, who's state will collapse until it is measured... It seems then that this argument can be continued indefinitely, without a point at which we can say that a measurement took place. We continue then without a recipe to determine when to use the Schrödinger equation and when the reduction

[5] We can think of an electron as being the particle to be used and the spin along a particular direction as the property to be measured (the *color* of the electron). Spin is an intrinsic form of angular momentum carried by quantum particles; it is a solely quantum-mechanical phenomenon with no counterpart in classical mechanics. For electrons, which are, so-called, spin one-halve 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.

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 devises 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 devises must be treated quantum-mechanically. Of course, by doing so, one allows for measurement devises to enter superpositions, and the standard way to match that, 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 vector is complete,
2. Quantum vectors 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 by negating (3) at many-world scenarios [Everett, 1957]. In order to negate (2), one needs to specify when is the Schrödinger equation interrupted.

The standard interpretation discussed above lands in this third category that negates (2). However, as we mentioned, it is unsatisfactory because it relies on the undefined notion of measurement in order state when the

Schrödinger equation is no longer valid. Nevertheless, one can take this third route and do better. One option, taken by objective collapse models such as GRW [Ghirardi et al., 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 possible way out of this situation, which has been proposed repeatedly throughout the years, is to invoke consciousness in order to break the above mentioned regress (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, a consciousness 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, on the one hand, Stapp [1993, 1996, 2005, 2006] has defended that the colapse of the wave function depends on consciousness; on the other, Hameroff and Penrose [1996] and more clearly Hameroff and Penrose [2013] deny that conscious observation causes quantum state reduction, and rather postulate an identity between the two phenomena (ibid. p.29). Finally, other quantum theories of consciousness that remain neutral on the relation between consciousness and reduction are not targeted by this paper.[6]

2 The Naïve Way to Try (and Fail) to Refute Quantum Theories of Consciousness

Whether consciousness is necessary or not 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 set up needed is not that complicated: one should simply seal in a box a quan-

[6] For 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, which initiates the chemical synapsis—without any apparent relation to the quantum reduction.

tum system and a measuring device[7] (MD), along with a mechanism that allows them to interact only at some given time in the future (let's say at noon).[8] MD is further equipped with a display that, as soon as a definite result is obtained, it shows both the result and the time at which the result 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 we find in it written a definite value recorded at noon, then it seems we can conclude that MD was enough to cause the collapse and so a consciousness is not required for it.

A similar experiment is proposed by Koch and Hepp [2006] in a recent article in *Nature*, attempting to dismiss quantum theories of consciousness in favor of neurobiological ones. Koch and Hepp make use of an example of binocular rivalry where one of the eyes is presented with a salient stimulus, like rapidly changing faces, whereas the quantum system is presented to the other. In these circumstances, 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 the famous Schrödinger's box with the live and dead cat, Koch and Hepp 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

[7] There 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 "conscious-free" system (for example a system that does not include a brain in Stapp's proposal or living cells containing microtubule-associated proteins in Hammerof and Penrose's one), with different states which are clearly distinguishable by a human being.

[8] If one is suspicious about the fact that the measurement occurs at an established time because it might introduce a way in which consciousness might be related to the measurement performed by MD, then one can complicate the set up *mutatis mutandi* by introducing a random number generator in such a way that when a certain sequence obtains, MD performs the measurement.

> 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. (p.612)

It seems that what Koch and Hepp are suggesting is nothing more than the arrangement we described at the beginning of this section. However, this kind of proposals are based on a naïve 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 experiment described above 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 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 thing 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 consciousness is or is not 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

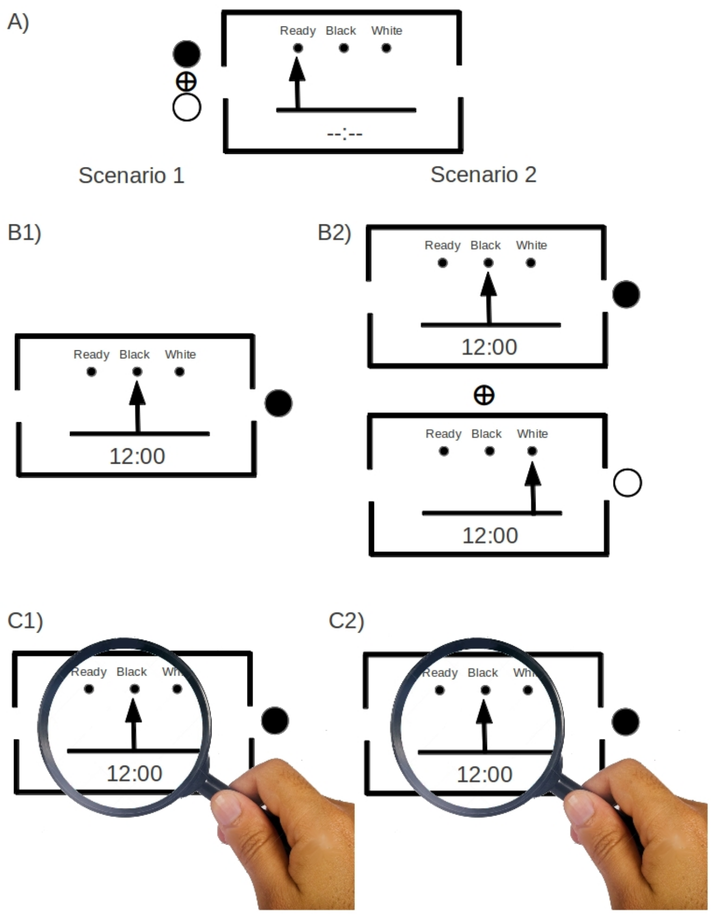

Fig. 1: An experiment often proposed to settle the discussion, which does not work. Scenario 1: consciousness not required for measurements; Scenario 2: consciousness required for measurements. A) Situation before measurement; B) Situation after measurement but before a conscious observer opens the box; and C) Situation after the box is opened. "$\oplus$" denotes a superposition.

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 scenario, during their interaction, the particle and MD will evolve according to the Schrödinger equation and the result will be a state of superposition between the two possible results of the measurement: one in which the needle indicates Black and one in which it indicates White (analogously 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 latter, the situation for the observer will be indistinguishable from scenario 1.

In reply to Kock and Hepp, Stapp [manuscript] 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[9] 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",

[9] Heisenberg, W. Physics and Philosophy. p. 54 (Harper, New York, 1958).

> 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 conscious-based from conscious-free interpretations of measurement using Schrödinger's standard thought experiment (i.e., essentially what we described above), but claim that it is possible to do so with a more complicated arrangement. In this regard, they propose an experiment 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" (ibid. 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, let's say, 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 set-up 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). Carperter and Anderson performed an experiment with this set-up 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 what is at issue is whether an interaction between consciousness and the device is required and 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.

In order to see in more detail that the result in Carpenter and Anderson [2006] is not valid, we will show that the predictions of a conscious-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 state*s. 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 such point, the system, from the S_2 perspective, will be modeled by a so-called *mixed state* which includes two elements of indeterminacy, one due to S_2's ignorance about the set-up chosen by S_1 and another one 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 measures, she collapses the state to the term which contains the message she observes. And, importantly, this happens even though she is unaware of its truthness or falseness, (this is, in fact, a well-known quantum phenomenon, present, for example, in the famous EPR thought experiments [Einstein et al., 1935], where local measurements collapse the whole state, even though part of it might be inaccessible to the measurer). As a result, one would expect, according to this theory, results identical to those observed in Carpenter and Anderson's experiment. In particular, according to conscious-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 quantum-free and quantum-based theories of measurement can be distinguished. In the next section we will argue that it is not and show an empirical way to discriminate, *pace* Stapp, scenario 1 from scenario 2 and thereby determine whether defenders of the quantum theories of consciousness under consideration make the right kind of predictions.

3 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 remembering that in quantum mechanics a superposition of, say, the states $|A\rangle$ and $|B\rangle$ cannot be interpreted as saying that the system is in either state $|A\rangle$ or state $|B\rangle$ but that we do not know which. That is because there are measurable properties possessed by the superposition that are not possessed by either $|A\rangle$ or $|B\rangle$ separately. Therefore, if we need to decide whether the state of a system is in the superposition $\alpha|A\rangle + \beta|B\rangle$ or, either $|A\rangle$ or $|B\rangle$, we can measure the system to see if such measurable properties of the superposition obtain or not. In more detail, we know that any quantum state possesses *with certainty* a specific property,[10] and so, if the state of a system is known, there is a property such that if it is measured, we are sure to obtain as a result a particular value that can be predicted with certainty. Therefore, in order to distinguish between $\alpha|A\rangle + \beta|B\rangle$, and a state in which either $|A\rangle$ or $|B\rangle$ is the case, we can measure such property, and if the state is the superposition one we will necessarily obtain the corresponding predictable value. If, on the other hand, the state is either $|A\rangle$ or $|B\rangle$, one will obtain different results.[11]

To see how all this works in more detail, let's 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 this state it will automatically collapse into either black or white. However, as we have just seen, there must be another property such that if we measure it, we will know with certainty whether the particle was in a superposition or not. Following Albert [1992]

[10] That is 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 vector $|\psi\rangle$ there are many observables for which such state is an eigenstate. A particularly simple observable that does the job is the projector $P_{|\psi\rangle} = |\psi\rangle\langle\psi|$ which has $|\psi\rangle$ as an eigenstate with eigenvalue 1 and any state orthogonal to $|\psi\rangle$ as an eigenstate with eigenvalue 0.

[11] Actually, even if the state is given by 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 and as soon as one obtains a value different than the one associated with the superposition, one can claim that the state was either $|A\rangle$ or $|B\rangle$ and not the superposition.

once more, we will call such property "*hardness*", and its two possible results "hard" and "soft".[12] 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 then we can conclude that the particle was not in a superposition. This, but applied to the pair particle-MD, instead of only to 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 "conscious-free" system, with states corresponing to different outcomes which are macroscopically distinct, i.e., clearly distinguishable by a human being. 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 of measuring different results (B2 in Fig. 2); that is, one in which the particle is black and MD displays such result and another one in which the particle is white and MD displays that result; 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

[12] If we continue playing along with the idea that our particle is an electron and that *color* is the spin along z, we can think of *hardness* as the spin along some direction different than z and of the value hard as spin up along that direction. In fact, all states of a spin one-half 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 such 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 x").

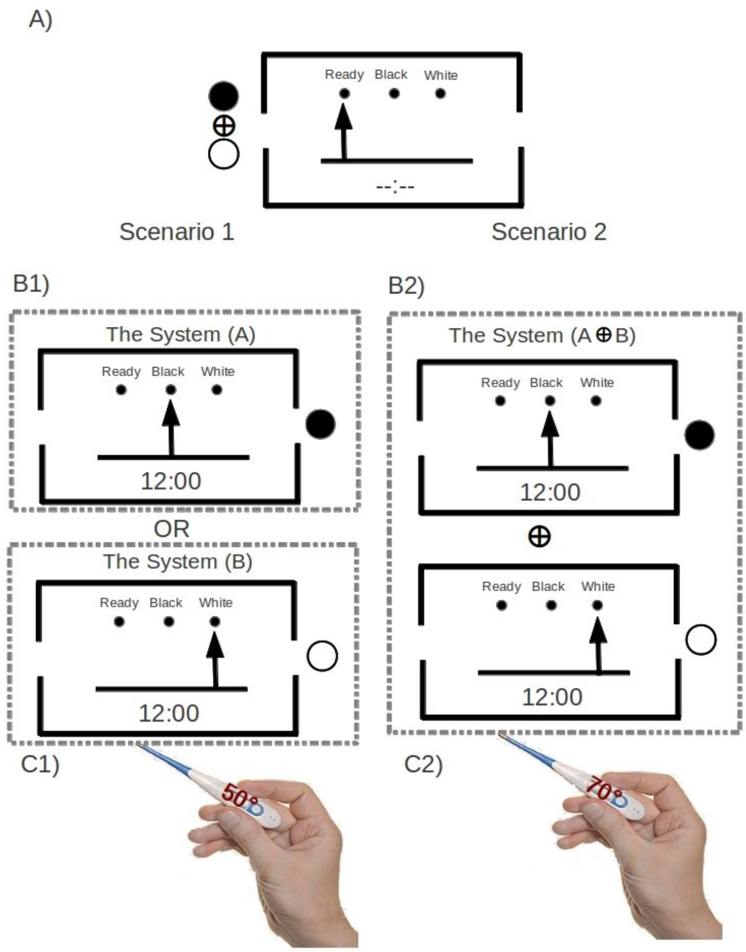

Fig. 2: An experiment to settle the discussion. Scenario 1: consciousness not required for measurements; Scenario 2: consciousness required for measurements. A) Situation before measurement; B) Situation after measurement but before a conscious observer measures *temperature*; and C) Situation after *temperature is measured*. "$\oplus$" denotes a superposition.

such result *or* one in which the *color* of the particle is white and MD displays that result (B1 in Fig. 2). Call these states $|A\rangle$ and $|B\rangle$ respectively. At such 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 there being 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 procedure we applied above to the particle.

The next step, then, is to bring in a conscious observer to measure TS.[13] 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 the superposition. Then, a more complicated property of TS will have to be measured; we will call such property "*temperature*" and denote by "70°" the value of *temperature* that characterizes the superposition, (i.e., the value we would expect to find with certainty if TS were in the state $|A \oplus B\rangle$). Therefore, 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 such measurement will, given what we explained above, 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°), then a collapse did occur when MD and the system interacted, meaning that consciousness is not required for collapses to occur. If, on the other hand, it is (the value of *temperature* is always found to be 70°), then a collapse did not occur when MD interacted with the system. In this case, a theory like the one proposed by Hameroff and Penrose [1996, 2013] would

[13] We remind the reader that our objective is to propose an experiment to discriminate between two types of *collapse theories*, those that involve consciousness and those that do not.

be falsified because it predicts that the reduction of the state of TS should happen long before the conscious observation does.[14] This is, of course, insufficient for showing that consciousness is required for the collapse of the wave function. 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 for thinking that it is something inherently human what 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 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 in the air surrounding it); and as soon as such an interaction occurs, the *temperature* measurement we propose stops being a reliable method to determine whether the state of TS is $|A \oplus B\rangle$ or not. That is due to the fact that, after an interaction with even a single air particle, it might no longer be possible to assign a pure quantum state to TS (just as

[14] In their reply to Koch and Hepp [2006], Hameroff and Penrose make explicit this commitment:

> 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 $\tau \approx \hbar/E_G$ 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. [Hameroff and Penrose, 2013, p. 29; our emphasis]

it is impossible to assign a pure quantum state to the original particle after its interaction with MS), in which case a measurement of *temperature* of 70° will no longer indicate that TS is in the state $|A \oplus B\rangle$. What we need, then, is to be able maintain, for a period of time sufficiently long to measure the *temperature,* a state like $|A \oplus B\rangle$.[15]

Difficult as that might seem, amazing advances in the construction and preservation of quantum superposition of distinct macroscopic states, such as the one of our proposal, have been achieved lately. For example, Friedman et al. [2000] presents experimental evidence that a superconducting quantum interference device (SQUID) can be maintained in a superposition of two macroscopically distinct magnetic-flux states. Moreover, Bruno et al. [2013], 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 MS in a measurement of a property, the *color*, of the single photon. That is, the millions of photons would play the role of the macroscopic MS, the single photon the role of the microscopic quantum system and together they would form TS which needs to be measured in order to asses if it indeed requires a conscious measurer in order to 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 one macroscopic measuring devise, not for all of them.

[15] At this point it is important to stress that the widespread believe 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, one must note that even if the reduced density matrix in question has the same form as an improper mixture, it does not follow that the physical situation of the subsystem is identical to that of the ensamble described by the identical improper mixture.

One might also object that, given that human brains are warm and wet, one must assing to them mixed states and not pure ones. Note however that, in order for our proposal to work, one needs to assing a pure state to TS but not to the conscious brain that measures it at the end. We conclude from all this that, in the foreseeable future, the proposed experiment could allow to either refute or confirm consciousness-based interpretations of quantum mechanics, not only in principle but in practice.

4 Conclusion

The idea of consciousness playing a key role in the determination of when measurements occur, and hence in controlling when collapses of quantum states happen, has been repeatedly offered throughout the years. In consequence, several attempts to dismiss this hypothesis have been presented. However, as we have shown, these attempts fail because they are based on a misunderstanding of the theoretical postulates involved. In this paper we have proposed instead a sound empirical way to determine whether consciousness is involved in measurements or whether collapses can happen independently of consciousness.

In particular, we have proposed an experiment such that, if it is found that the interaction of a system in a superposition state with a measurement device results in a determinate state, then one can conclude that consciousness is not required for collapses to occur. If so, one important motivations for looking into quantum mechanics for a theory of consciousness would be lost, 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 possess with certainty, then this would give us good reasons to think that consciousness is required for quantum measurements, urging us to look further into the quantum realm in order to construct a theory of consciousness.[16]

References

David Z. Albert. *Quantum Mechanics and Experience.* Harvard University Press, 1992.

Katalin Balog. Phenomenal concepts. In Brian McLaughlin, Ansgar Beckermann, and Sven Walter, editors, *Oxford Handbook in the Philosophy of Mind.* Oxford University Press, 2009.

F. Beck. Quantum brain dynamics and consciousness. In P. van Loocke, editor, *The Physical Nature of Consciousness.* Amsterdam: Benjamins, 2001.

F. Beck and J. Eccles. Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences of the USA*, 89:11357–11361, 1992.

D. Bohm. A suggested interpretation of the quantum theory in terms of hidden variables. i & ii. *Physical Review, 85:166*, 1952.

N. Bruno, A. Martin, P. Sekatski, N. Sangouard, R. T. Thew, and N. Gisin. Displacement of entanglement back and forth between the micro and macro domains. *Nature Physics*, 9:545–548, 2013.

R. H. S. Carpenter and Andrew J. Anderson. The death of schoedingers cat and of consciousness based quantum wave-function collapse. *annales de la Foundation Loisi de Broglie*, 33:45–52, 2006.

David Chalmers. *The Character of Consciousness.* Oxford University Press, 2010.

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

David J. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, USA, 1 edition, November 1996. ISBN 0195117891.

David J Chalmers. Consciousness and its place in nature. In Stephen P Stich and Ted A Warfield, editors, *Blackwell Guide to the Philosophy of Mind*. Blackwell, 2003.

David J Chalmers. Phenomenal concepts and the explanatory gap. In Torin Alter and Sven Walter, editors, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press, 2007.

David J Chalmers. The Two-Dimensional argument against materialism. In Brian P McLaughlin and Sven Walter, editors, *Oxford Handbook to the Philosophy of Mind*. Oxford University Press, 2009.

A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777, 1935.

Hugh Everett. "relative state" formulation of quantum mechanics. *Review of Modern Physics*, 29:454, 1957.

Richard Feynman. *The Character of Physical Law*. Modern Library, 1994.

Jonathan R. Friedman, Vijay Patel, W. Chen, S. K. Tolpygo, and J. E. Lukens. Quantum superposition of distinct macroscopic states. *Nature*, 406:43–46, 2000.

G.C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review, D34:470*, 1986.

S. Hameroff and R. Penrose. Conscious events as orchestrated spacetime selections. *Journal of Consciousness Studies*, 3:36–53, 1996.

S. Hameroff and R. Penrose. Consciousness in the universe: A review of the Orch OR theory. *Phys Life Rev*, 2013.

Christopher S Hill and Brian P Mclaughlin. There are fewer things in reality than are dreamt of in chalmers's philosophy. *Philosophy and Phenomenological Research*, 59(2):445–454, 1999.

C. Koch and K. Hepp. Quantum mechanics in the brain. *Nature*, 440:611–612, 2006.

Joseph Levine. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(October):354–61, 1983.

Brian Loar. Phenomenal states. *Philosophical Perspectives*, 4:81–108, 1990.

A. I. Lvovsky, R. Ghobadi, A.. Chandra, A. S. Prasad, and C. Simon. Observation of micro-macro entanglement of light. *Nature Physics*, 9:541–544, 2013.

Tim Maudlin. Three measurement problems. *Topoi*, 1995.

Elias Okon. El problema de la medicion en mecanica cuantica. *Revista Mexicana de Fisica*, 60:130–140, 2014.

H. Primas. Hidden determinism, probability, and time s arrow. In H. Atmanspacher and R.C. Bishop, editors, *Between Chance and Choice*. Exeter: Imprint Academic, 2002.

H. Stapp. Quantum interactive dualism: An alternative to materialism. *Journal of Consciousness Studies*, 12:43–58, 2005.

H. Stapp. Reply to koch-hepp essay in nature. manuscript.

H.P. Stapp. *Mind, Matter and Quantum Mechanics*. Springer, 1993.

H.P. Stapp. Attention, intention, and will in quantum physics. *Journal of Consciousness Studies*, 6:143–164, 1996.

H.P. Stapp. Clarifications and specifications in conversation with harald atmanspacher. *Journal of Consciousness Studies*, 13:67–85, 2006.

Michael Tye. Phenomenal consciousness: The explanatory gap as a cognitive illusion. *Mind*, 108(432):705–25, 1999.

D. Wallace. The quantum measurement problem: State of play. In Dean Rickles, editor, *The Ashgate Companion to Contemporary Philosophy of Physics*. Ashgate, 2008.