2

The Subjective-Objective Collapse Model

Virtues and Challenges

Elias Okon and Miguel Ángel Sebastián

1. Introduction

Quantum mechanics certainly is one of the most successful physical theories ever constructed. However, its standard interpretation suffers from a grave conceptual issue known as the *measurement problem*. Such a problem consists of the fact that the framework crucially depends on notions such as *measurement* or *observer*, but such notions are never formally defined within the theory. A direct consequence of this is an ambiguity regarding where in the causal chain between the interaction of the system under study with a measuring device and our "subjective perception" is the proverbial collapse of the wave function to be placed.

The idea that consciousness causes the collapse of the wave function was first suggested in London and Bauer (1939) and then developed in Wigner (1967). The latter proposed that consciousness, unlike other properties, does not admit superpositions. This, however, seems to suggest that consciousness is not a physical property or, at least, not a standard one. More recently, other works have explored similar ideas—see, e.g., Chalmers and McQueen (MS); Stapp (2007). The problem, though, is that this avenue of research has never been popular among physicists, as it seems to commit us to some form of dualism, in which the interaction between consciousness and the physical remains obscure—if not downright inconsistent with certain basic principles of physics, such as conservation of energy (Averill and Keating, 1981; Larmer, 1986).

In Okon and Sebastián (2018), we presented a physicalist-friendly model in which, in a well-defined sense, consciousness "causes" the collapse of the wave function. The model consists of an objective collapse scheme, where the collapse operator is associated with consciousness as a physical property—

Elias Okon and Miguel Ángel Sebastián, *The Subjective-Objective Collapse Model: Virtues and Challenges*
In: *Consciousness and Quantum Mechanics*. Edited by: Shan Gao, Oxford University Press.
© Oxford University Press 2022. DOI: 10.1093/oso/9780197501665.003.0003

or with a physical property that perfectly correlates with consciousness. We call it the "Subjective-Objective Collapse" (SOC) model. SOC is such that superpositions of conscious states are dynamically suppressed in a way that is fully compatible with our subjective experience. As such, it opens up new lines of research, both in fundamental physics and consciousness studies.

In this chapter we evaluate the virtues of the model and analyze some possible objections and challenges. For this purpose, the chapter is organized as follows. Section 2 presents the measurement problem and the solution offered by the SOC model. In section 3 we question the compatibility of the model with certain theories of consciousness and, in particular, with the idea that consciousness might be multiply realizable—an idea that is, *prima facie*, in tension with the model. Finally, in section 4 we examine important aspects of the SOC model in light of considerations in the field of quantum foundations.

2. The Measurement Problem and the Subjective-Objective Collapse Model

2.1 The Measurement Problem

The measurement problem can be stated as the fact that standard quantum mechanics crucially depends on notions such as measurement or observer, even though such notions are never formally defined within the theory—see, e.g., Bell (1990). In particular, measurements or observers are employed to decide when the indeterministic collapse process is supposed to interrupt the deterministic Schrödinger evolution. The problem, of course, is that, without a firm grasp of such higher-level notions, one ends up with undesirable vagueness in an otherwise fantastically successful theory.

To explore the issue, let us suppose, for now, that everything evolves according to the Schrödinger equation *at all times*. With this in mind, let us consider an apparatus with a ready state $|R\rangle_M$ that, when fed with a spin-1/2 particle, behaves as follows:

$$|R\rangle_M|+\rangle_p \xrightarrow{\text{Schrödinger}} |+\rangle_M|+\rangle_p \quad \text{and} \quad |R\rangle_M|-\rangle_p \xrightarrow{\text{Schrödinger}} |-\rangle_M|-\rangle_p \quad (1)$$

where $|+\rangle_M$ and $|-\rangle_M$ are states of the apparatus in which it displays spin-up and spin-down as the result of the experiment. That is, the apparatus correctly measures the spin—say, along z—of the particle. Now, what happens if the apparatus is fed with a particle in a *superposition* of $|+\rangle_p$ and $|-\rangle_p$? Well, the linearity of the Schrödinger equation leads to

$$|R\rangle_M\{\alpha|+\rangle_p + \beta|-\rangle_p\} \xrightarrow{\text{Schrödinger}} \alpha|+\rangle_M|+\rangle_p + \beta|-\rangle_M|-\rangle_p \tag{2}$$

We see that the apparatus ends up in a superposition of displaying spin-up and spin-down, which seems odd, to say the least.

To push the matter further, consider the introduction of an observer in a state $|R\rangle_O$, in which she is ready to see the display of the apparatus during the measurement. Once again, the linearity of the Schrodinger equation leads to a superposition, but this time of the observer perceiving the apparatus displaying spin-up and spin-down:

$$|R\rangle_O|R\rangle_M\{\alpha|+\rangle_p + \beta|-\rangle_p\} \xrightarrow{\text{Schrödinger}} \alpha|+\rangle_O|+\rangle_M|+\rangle_p + \beta|-\rangle_O|-\rangle_M|-\rangle_p \tag{3}$$

where $|+\rangle_O$ and $|-\rangle_O$ are the states in which the observer perceives that the apparatus displays spin-up and spin-down. The problem, of course, is that the final state above describes a *superposition of perceptions*, which is not what we seem to experience when we perform this type of experiments. Of course, according to standard quantum mechanics, the final state in equation (3) is not the final state of the system because the collapse postulate has been left out of this discussion, but we have already mentioned that the introduction of such a postulate in the standard framework is problematic. So what could be done in order to solve the problem. Well, a satisfactory solution to the measurement problem consists of a formalism which:

1. Is fully formulated in precise, mathematical terms (with notions such as *measurement*, *observer* or *macroscopic* not being part of the fundamental language of the theory).
2. Reproduces the empirical success of standard quantum mechanics at the microscopic level.
3. Explains why certain macroscopic superpositions formally allowed by the theory never seem to occur.[1]

[1] Regarding this last point, it is important to stress an often overlooked distinction between a *superposition of perceptions* and a *perception of a superposition*. That is, between:

At least three strategies have been suggested in order to achieve this. The first one, introduced by Everett (1957), consists of attempting to read the final state in equation (3) not as a state in which the observer does not have a well-defined perception, but one in which the observer simultaneously, but *independently*, has both perceptions. Alternatively, one could avoid having to interpret the final state in equation (3) as one in which the observer does not have a well-defined perception by adding extra elements to the picture—such as Bohmian particles—that determine which of the two terms of the superposition actually obtains in the world—see Goldstein (2013). Finally, one could modify the dynamics in such a way that the actual final state of the system is not a macroscopic superposition, but one of the alternatives; this is the option that SOC explores.

2.2 Objective Collapse Models

The *objective collapse* (or *dynamical reduction*) program aims at constructing a modified quantum dynamics that merges the standard unitary evolution and the collapse mechanism. The idea is to add non-linear, stochastic terms to the Schrödinger equation that suppress unwanted macroscopic superpositions.

The simplest collapse model is know as GRW (Ghirardi et al. (1986)). In it, all elementary particles are postulated to suffer spontaneous localization processes around positions selected according to a probability distribution that approximates the Born rule. Since the frequency of collapses is proportional to the number of particles, macroscopic objects are very likely to undergo them, even if the collapse rate at the level of elementary particles is extremely small. Additionally, due to the fact that the collapse of one of the particles of a macroscopic superposition—such as that in the final state in equation (3)—is sufficient for the whole body to get localized, GRW

- A superposition of incompatible perceptions, such as the final state in equation (3): $\alpha|+\rangle_O|+\rangle_M|+\rangle_p + \beta|-\rangle_O|-\rangle_M|-\rangle_p$.
- A well-defined perception of a macroscopic system in a superposition of different positions: $|S\rangle_O\{\alpha|+\rangle_M|+\rangle_p + \beta|-\rangle_M|-\rangle_p\} = |S\rangle_O|S\rangle_{M+p}$ (where $|S\rangle_{M+p} \equiv \alpha|+\rangle_M|+\rangle_p + \beta|-\rangle_M|-\rangle_p$ and $|S\rangle_O$ corresponds to a state in which the observer experiences a well-defined perception of the measurement apparatus displaying a superposition of spin-up and spin-down).

We have already explained how the first case may arise and why it represents a problem. Regarding the second, from the *quantum* point of view, both $|S\rangle_{M+p}$ and $|S\rangle_O$ are states which are on a par with states such as $|+\rangle_O$ or $|+\rangle_M|+\rangle_p$, and the only reason they seems strange is because we in fact do not experience them. The key question, of course, is why.

guaranties superpositions of well-localized macroscopic states to die quickly, and in accordance with the statistics of standard quantum mechanics.

In the objective collapse model known as Continuous Spontaneous Localization, or CSL Pearle (1989): the discontinuous localization events of GRW are replaced by a continuous, stochastic process. In more detail, in CSL, particular non-linear, stochastic terms, which are designed to drive any initial wave function into one of the eigenstate of a, so-called, collapse operator, are added to the Schrödinger equation. In the simplest case, the CSL equation is such that its solutions are given by

$$|\psi(t)\rangle_B = e^{-\left\{iH+\frac{1}{4\lambda t}[B(t)-2\lambda\hat{A}]^2\right\}}|\psi(0)\rangle \tag{4}$$

where $\hat{A}$ is the, so-called, collapse operator and $B(t)$ is a classical Brownian function, stochastically chosen with a probability density given by

$$\mathcal{P}_t\{B\} = {}_B\langle\psi(t)|\psi(t)\rangle_B \tag{5}$$

To see why this achieves what it is supposed to, for simplicity we take $H = 0$ and expand $|\psi(0)\rangle$ in terms of eigenstates of $\hat{A}$:

$$|\psi(0)\rangle = \sum_i c_i|a_i\rangle \tag{6}$$

which leads to

$$|\psi(t)\rangle_B = \sum_i c_i e^{-\frac{1}{4\lambda t}[B(t)-2\lambda t a_i]^2}|a_i\rangle \tag{7}$$

and

$$\mathcal{P}_t\{B\} = \sum_i e^{-\frac{1}{2\lambda}[B(t)-2\lambda t a_i]^2}|c_i|^2 \tag{8}$$

Since the last equation implies that the most probable $B(t)$'s to occur are $B(t) \approx 2\lambda t a_j$, with probabilities $|c_i|^2$, we finally obtain

$$|\psi(t)\rangle_B \approx c_j|a_j\rangle + \sum_{i\neq j} e^{-2\lambda t[a_i-a_j]^2}|a_i\rangle \xrightarrow{t\to\infty} c_j|a_j\rangle \tag{9}$$

which means that, as $t \to \infty$, the CSL dynamics drives the state of the system into the jth eigenstate of the operator $\hat{A}$, with probability $|c_j|^2$; i.e., the CSL dynamics includes, along with the standard Schrödinger evolution, a "measurement" of the observable represented by the collapse operator.

In standard CSL models, the collapse operator $\hat{A}$ is chosen to be constructed out of the position operator. This is so because such a choice directly induces the desired suppression of superpositions of macroscopic objects at different locations. In fact, it has even been argued that a choice of this sort is the only alternative that could work—see Bassi and Ghirardi (2003). We showed in Okon and Sebastián (2018) that a very different choice for a collapse operator can also lead to a solution of the measurement problem. The point was that, in order to explain why superpositions of macroscopic objects are never actually perceived, (at least) two options present themselves: one can construct models in which such macroscopic superpositions never occur—as in standard collapse theories—or one can maintain that, although such superpositions do occur, we never encounter them because they collapse a soon as we observe them. We exploit this second group of alternatives by developing a CSL model in which the collapse operator is associated with consciousness: the SOC model.

2.3 The Subjective-Objective Collapse Model

The idea that consciousness causes the collapse of the wave function has never been popular, mainly because it appears to be in tension with physicalism. Moreover, if, as Wigner, one assumes that consciousness never superposes, then systems would never get entangled with conscious beings and it would not be clear how a consciousness-based collapse would help explaining why we do not observe macroscopic superpositions.

Chalmers and McQueen (MS) have proposed, as a solution to the measurement problem, the introduction of what they call m-properties or superposition-resistant observables, whose superpositions are postulated to be either forbidden—which we have just seen that is problematic—or to be "unstable" or "more likely to collapse." Moreover, they suggest that some physical correlate of consciousness, such as integrated information, could be a superposition-resistant observable. What is missed is a concrete dynamical model that accommodates these suggestive ideas. For this to work, one needs to make sure that these superpositions quickly evolve into states of well-defined consciousness in such a way that we would fail to notice these transitions in our experience.[2] The SOC model solves this problem. In a few words, it consists of a CSL model in which the collapse operator depends on consciousness. By doing so, we arrive at a model in which, as has been suggested throughout the years, consciousness plays a role in the collapse of the wave function. The advantage of this proposal, of course, is that we incorporate consciousness into quantum theory in a perfectly well-defined way, both mathematically and conceptually, and in a way which is fully compatible with the truth of physicalism if consciousness is physical.

The first thing we need in order to build our model is a physical property upon which consciousness is supposed to depend; let's call this property Φ. What is needed next is the construction of a quantum version of Φ. For this, first we note that, according to quantum theory, to every well-defined property corresponds a Hermitian operator. If Φ is a well-defined property, it follows that there must be a corresponding operator $\hat{\Phi}$. The proposal, then, is that only states with well-defined values of $\hat{\Phi}$ correspond to conscious states, i.e., only eigenstates of $\hat{\Phi}$ are conscious.

With $\hat{\Phi}$ in our hands, we can finally introduce our model. For a given initial state $|\psi(0)\rangle$, the SOC model has as solutions

$$|\psi(t)\rangle_B = e^{-\left\{iH+\frac{1}{4\lambda t}\left[B(t)-2\lambda\hat{\Phi}\right]^2\right\}}|\psi(0)\rangle \tag{10}$$

with $B(t)$ a classical Brownian motion function selected randomly with probability density

$$\mathcal{P}_t\{B\} = {}_B\langle\psi(t)|\psi(t)\rangle_B \tag{11}$$

Therefore, according to what we said about the CSL dynamics above, the SOC model is such that it drives any initial state into an eigenstate of the $\hat{\Phi}$ operator. If, as we proposed above, eigenstates of $\hat{\Phi}$ are indeed related to conscious states—by measuring consciousness or a property that perfectly correlates with it in the actual world—then, in the same way that standard CSL quickly destroys superpositions of macroscopic objects localized at

[2] One might be puzzled at first sight by the idea of a conscious state we fail to notice. It is nonetheless easy to make sense of it by means of the conceptual distinction in philosophy of mind between phenomenal consciousness—the experience we have—and access consciousness (Block, 2002)—what we come to notice. There is even empirical evidence that supports the claim that, in cognitive systems like ours, the mechanisms underlying these faculties are in fact segregated (Block, 2011, 2014; Sebastián, 2014).

different places, the above dynamics quickly kills superpositions of incompatible conscious states, leading to states of well-defined consciousness.

Looking back at the list of requirement for a satisfactory solution to the measurement problem presented above in section 2, we see that the SOC model seems promising. To begin with, it is fully formulated in precise, mathematical terms and, as long as the λ parameter in equation (10) is small, it reproduces the empirical success of standard quantum mechanics at the microscopic level—see section 4.1. Regarding an explanation of why certain macroscopic superpositions allowed by the standard theory never seem to occur, the model clearly takes care of the complication by not letting those states last for long.[3]

It is important to note that the standard choice for a collapse operator in CSL, in terms of position, is of course well justified by observations, but lacks an explanation or an independent motivation. In SOC, in contrast, the fact that we never observe superpositions in the position of macroscopic objects is simply a contingent fact, derived from the cognitive architecture that happens to give rise to consciousness in our case.

Despite these virtues, SOC faces its own devils. In the rest of the chapter we deal with some of them. In particular in section 3, we analyze the adequacy of the SOC model in light of different theories of consciousness present in the literature, and in in section 4, we analyze the adequacy of the SOC model as a quantum theory.

3. The SOC Model, Theories of Consciousness and Multiple Realizability

According to the SOC model, superpositions of conscious states are quickly suppressed, which in turn explains why we fail to observe macroscopically superposed objects. The SOC model achieves this by driving any initial state into an eigenstate of an operator that measures consciousness. In the original

[3] As we mentioned in footnote 1, it is very important to highlight an often overlooked distinction between two different scenarios: i) superpositions of incompatible perceptions and ii) well-defined perceptions of a macroscopic system in a superposition of different positions. Our model straightforwardly takes care of the first by not letting those states last for long. However, models in which the collapse of the wave function depends on consciousness, admit the possibility of macroscopic asstates which correspond to well-defined perceptions of a macroscopic system in a superposition of different positions are *not* suppressed by the model. Therefore, a way of restricting access to those states is required. We deal with this case in detail in Okon and Sebastián (2018).

proposal, such an operator ($\hat{\Phi}$) measures integration of information which, according to IIT, is a measurement of consciousness. However, all the SOC model really requires is for there to be a physical property that corresponds to consciousness; that is, for consciousness to be identical with such a property or, at least, for consciousness to perfectly correlate with it in the actual world. The SOC model is then neutral on the fundamental nature of consciousness. If the nature of consciousness is not physical, then SOC can still offer a dynamics that accommodates the idea that "consciousness collapses the wave function" insofar as there is a physical property that perfectly aligns with it in our world. More interestingly, the SOC model can provide such a dynamics even if consciousness is indeed a physical property. Unfortunately, it is not obvious that the SOC model is compatible with most theories of consciousness. In particular, it does not seem to be compatible with theories that endorse the possibility of very different physical set-ups underlying conscious states. To see why, we need to get some clarity regarding the notion of a "physical property."

According to standard quantum mechanics, there is a one-to-one correspondence between *physical properties* and Hermitian operators. This technical use of the notion of a physical property departs from common sense or ordinary use. Although, in the absence of an exhaustive definition, we might disagree on what counts as a physical property in such an ordinary sense—and hence disagree on whether, e.g., the property of *being a human* shall count as one of those—a clear case thereof suffices to illustrate the difference. *Being a table* is definitely a physical property in the ordinary common sense, but not obviously so in the required technical sense because one might reasonably think that there is no Hermitian operator that corresponds to such a property. The reason is that there are many different ways in which tables can be physically realized, and hence no underlying physical property in the required technical sense that all, and only, tables share.[4]

The problem for the SOC model is that it does not seem to be compatible with theories that endorse the multiple realizability of conscious states. Is there any reason to believe that, in the actual world, consciousness can be multiply realized? Yes, there is. Chalmers, e.g., has argued that two systems with a sufficiently fine-grained functional organization—to fix the mechanisms responsible for the production of behavior, and to fix behavioral

[4] From now on, unless otherwise indicated, we restrict our use of of the term "physical property" to the technical sense.

dispositions (Chalmers, 2010)—will be equally conscious and enjoy the same kind of experiences. Accordingly, what matters for consciousness is a certain—sufficiently fine-grained—functional organization, and once this functional organization is satisfied, we can abstract from its particular realization. As Chalmers presents the idea:

> [W]hat matters for the emergence of experience is not the specific physical makeup of a system but the abstract pattern of causal interaction between its components. (*ibid*., p. 24)

Imagine that the required sufficiently fine-grained functional organization at which behavioral dispositions are fixed is that of neural networks. Neurons in our brain have a certain biochemical composition, but this composition is irrelevant if Chalmers is right. If a silicon chip can satisfy the same pattern of causal interaction as a neuron, then it would be possible to replace our neurons by those silicon chips without a change in the required functional organization and, therefore, in our conscious experience. This is what Chalmers calls the principle of *organizational invariantism*. Although this idea has not gone without controversy, Chalmers (1996, ch. 7) provides two convincing and complementary arguments in its favor: the fading/absent qualia and the dancing qualia arguments. Roughly, the arguments go as follows.

In the fading/absent qualia argument, we are asked to consider someone having, say, an experience of pain and, for the sake of a *reductio*, the possibility of a functional duplicate with a "brain" made out of silicon chips, which does not experience the pain at all. As the two systems have the same functional organization, we can imagine gradually transforming one into the other by replacing neurons by the corresponding silicon chips without changing the functional set-up. Two things might happen during the transformation: either the replacement of a single neuron switches off consciousness or the experience fades slowly along the process with every replacement. None of the alternatives is plausible, or so argues Chalmers. The first one because it requires that "there would be brute discontinuities in the laws of nature unlike those we find anywhere else" (*ibid*., p. 238). The second one because it would require that a system, whose cognitive processes are not malfunctioning and that is conscious, be systematically wrong about its own experience, complaining about its horrible pain while it is merely having a really mild one.

In the dancing qualia argument, we also consider a transformation process from a system with a neuronal brain to a system with a silicon brain. However, in this case, we assume, again for the sake of a *reductio*, that they have different experiences; e.g., that after the replacement the subject has an experience as of blue while looking at a red apple. If the principle of organizational invariantism were false, when we switch from the neuronal brain to the silicon one, the subject's experience would change from an experience as of red to an experience as of blue, but such a change in experience would go unnoticed for her. What is more, we can imagine flipping the switch back and forth so that "the red and blue experiences 'dance' before [S's] eyes" (Chalmers, 1996, p. 253), but S does not notice any change. This does not seem plausible according to Chalmers. The fading and the dancing qualia arguments provide good support for thinking that consciousness is multiply realizable and hence, as we have suggested, incompatible with the SOC model.

Most (philosophical) theories of consciousness accommodate multiple realizability, and hence, seem to be in tension with the SOC model. Let us briefly discuss the details of the relation between different families of theories of consciousness and multiple realizability in some detail.

Dualist positions—which deny that consciousness is ontologically different from, and does not metaphysically depend on, physical properties—are perfectly compatible with the SOC model insofar as there is a physical property that, in the actual world, perfectly aligns with consciousness. But dualists can also deny that consciousness perfectly correlates with a physical property, endorsing the principle of organizational invariantism (Chalmers, 1996) and in tension with the SOC model.

Panpsychists defend the idea that our conscious experience depends upon the very nature of the most fundamental particles. Multiple realizability seems to be in tension with panpsychism because different systems can satisfy the same sufficiently fine-grained organization and yet differ significantly at the fundamental level (Sebastián, 2015). This makes panpsychism *prima facie* perfectly compatible with the SOC model.

Materialist theories can endorse an identity theory that would be compatible with the SOC model. However, most materialist theories accommodate multiple realizability.

Functionalists commonly hold that conscious states are those that satisfy a certain causal role to be determined either *a priori*, as in Lewis (1978);

Dennett (1991), or *a posteriori*, as in Baars (1988); Prinz (2012).[5] If such a causal role can be multiply realized, then functionalism seems also in tension with SOC as it is reasonable to assume that there is no physical property commonly instantiated by all the possible realizers.

Representationalists claim that conscious states are representational states, i.e., states that have adequacy conditions. Our conscious states represent the world and ourselves as being a certain way and we say that they are adequate or correct depending on whether the world is that way. For example, the conscious experience we have when looking outside the window can be evaluated as correct or incorrect depending on whether it is the result of the interaction with the environment or of the consumption of certain toxic substance.[6] The consistency of representationalism both with materialism and with functionalism depends upon a theory of mental content that makes it explicit what it takes for a system to be in the required representational state. Those that endorse a naturalistic approach typically and very roughly maintain that representational states are those that have the teleological function—one that determines what a trait should do rather than what it actually does—of carrying information about its object. For example, oversimplifying, one might think that the teleological function of the traits of biological entities like us depend on natural selection, and hence that a state to carry information as a state requires two things, i) there being reliable correlations between the state and the object, and ii) there being a certain kind of sender-receiver structure that allows us to exploit such correlations. This does require a common physical property, as the SOC model seems to assume.

Most theories of consciousness are consistent with multiple realizability. If any of those theories is correct, then it seems reasonable to think that there is no unique physical property that conscious states share, and hence that we cannot construct the operator that the SOC model requires. In reply, one could think of consciousness as being identical (or perfectly correlating), rather than with a physical property, with the property that results from the disjunction of all the possible realizers—the property of being this or that physical property. This in turn would require for the model to have as many equations as physical properties related to consciousness the realizers might

[5] Enactivist views can be seen as complex functionalist views (Hutto and Myin, 2013; Noë, 2005).

[6] Representationalists disagree on the representational properties of the corresponding state, and can thereby be taxonomized depending on the alleged content of experience.

have. This would make the model not only unappealing, but also hardly palatable. At any rate, we are going to argue that multiple realizability is not incompatible with the SOC model, as it *prima facie* seems.

According to the dynamic equation that the SOC model proposes, equation (10), superpositions of eigenstates of property Φ, associated with the Hermitian operator $\hat{\Phi}$, are quickly suppressed. If Φ measures consciousness, then we have an explanation of why we fail to observe superpositions of macroscopic objects. However, for the model to be satisfactory, we need to be able, in principle, to define $\hat{\Phi}$. We typically think of the properties associated with an Hermitian operator in a suitable Hilbert space as the fundamental ones that physics postulates. And one might reasonably think that, if consciousness is multiply realizable, there is no fundamental property that distinguishes conscious and unconscious states, in the very same sense as there is no distinctive fundamental property shared by, and only by, all tables. If we can only define Hermitian operators over fundamental properties, then it is not possible to construct the operator $\hat{\Phi}$ that the model requires.

Fortunately for the SOC model, the latter assumption is wrong: we can construct Hermitian operators that correspond to properties that are not fundamental. Conforming to any theory of consciousness that we would want to consider, there are going to be arranged collections of fundamental particles that correspond to conscious states and collections that do not—and maybe, collections such that it is indeterminate whether they correspond to conscious states or not. If there is a fundamental property that individuates the set of all conscious states, then we can uncontroversially construct $\hat{\Phi}$. If, on the other hand, there is none, then consciousness cannot be reduced to a set of fundamental properties—as the property of *being a table* arguably cannot either. This, however, does not mean that we cannot construct $\hat{\Phi}$, where Φ is either the property of being conscious or a perfect correlate. Given the adequate Hilbert space, all that we need to do is to consider all states that are conscious and note that all those states, being microscopically distinguishable, must be represented by vectors that are orthogonal among themselves. Then, since for every set of orthogonal vectors there is a Hermitian operator that has the vectors of the set as eigenstates, there must be a Hermitian operator that has all conscious states as eigenstates. That operator is the $\hat{\Phi}$ we were looking for. This is a perfectly legitimate Hermitian operator, which simply might not correspond to any fundamental property, but that does correspond to the property of consciousness. This is all that is required, so the SOC model, after all, is compatible with theories of consciousness that contemplate multiply realizability—such as functionalism or representationalism.

4. The SOC Model as a Quantum Theory

In this section we examine some aspects of the SOC models in light of considerations in the field of quantum foundations. First we explore the issue of fixing the value of the collapse parameter and then we enquire about the physical interpretation of the formalism.

4.1 Choosing the Value of the Collapse Parameter

The value of the collapse parameter λ in standard collapse models has to satisfy competing constraints. On the one hand, λ cannot be too large because microscopic phenomena, which we know are well-described by a purely unitary evolution, would get disturbed. Moreover, a large λ would lead to a type of quantum Zeno effect (QZE) [see Dagasperis et al. (1974)], in which the collapse terms dominate and eigenstates of the collapse operator suffer recurring collapses and effectively freeze. On the other hand, if λ is too small, these models would not achieve their purpose of suppressing undesirable macroscopic superpositions. Of course, one can allow for these macroscopic superpositions to persist for some time, but one needs to make sure for them to quickly die out before we are able to notice them. The beauty of these collapse models is that there are values for λ that provide empirically successful models of the world around us (see Adler (2007); Feldmann and Tumulka (2012); Bassi *et al.* (2003)).

One could worry that our model could run into trouble with QZE. If present, the effect would mean that conscious states would not be able to change, contradicting our experience. The situation, however, is perfectly analogous with standard, position-based, CSL models which, as we just mentioned, are not affected by QZE. That is, in the same way that standard CSL does not imply that the position of macroscopic objects would not change, our model does not imply eigenstates of Φ to freeze. This is so because the actual CSL evolution is always dictated by the interplay between the Hamiltonian and the collapse terms, so the actual evolution of a system involves a competition between the two. Of course, the result of this struggle is to be decided by the strength of the collapse terms, which is determined by the parameter λ, so the key question is if there is a possible value for λ in our model that avoids these problems and yields empirically successful predictions.

In our model, as in standard CSL, λ is a free parameter, a new constant of nature if you will, that controls the strength of the collapse terms. Its value must then be chosen, if possible, to make sure the model is empirically adequate. In particular, λ must be chosen to avoid, on one hand, the QZE, and, on the other, perceptible Schrödinger cat states. But one might worry, though, that this may not be possible for our model.

To begin with, we notice that in GRW, λ is taken to be a very small number ($\approx 10^{-16}$ s^{-1}). However, the *effective* collapse rate of a system is given by $N\lambda$ where N is the number of particles of the system which are entangled and in a superposition of different positions. A macroscopic system in such a state of superposition is then extremely likely to undergo a collapse very soon, and if such collapses would continue at such a rate, then it would suffer from a QZE-type issue and practically freeze. This does not happen, however, because an initial collapse destroys the superposition, dramatically lowering the number of particles entangled, and thus the effective collapse rate. It is not clear, though, that a mechanism of this sort is present in our case. The issue described above contains in fact two related worries, one potentially afflicting CSL in general, and one specific to our model:

1. Unlike GRW, our model does not seem to contain a mechanism that suppresses collapses once an eigenstate (or something close to it) is reached; this could lead to a QZE-type problem.
2. Unlike GRW and standard CSL, the collapse rate in our model does not seem to scale with the number of particles.

Regarding 1, in spite of appearances to the contrary, the CSL model we employ does contain a mechanism that suppresses the collapse terms once the state is close to an eigenstate. This is not obvious from the CSL equation described above, which does not preserve the norm of the state (note the c_j multiplying the eigenstate at the end of equation (9)). However, it becomes clear if one writes an equation for a physical state that remains normalized during the evolution process. By doing so, one notices that the collapse terms are a function of the collapse operator *minus* its expectation value (see equaiton 3.13 in Pearle (1990)). Since, acting on an eigenstate, such a subtraction is zero, the collapse mechanism is, as desired, ineffective on such states.

Regarding 2, in GRW one postulates a collapse process at the micro level with fix rate λ_{GRW}, and depends on the state of the whole system having special correlations for the collapse to be significant at the macro level. As a result, one ends up with an effective collapse rate at the macro level that depends on the details of the macro state. In CSL, on the other hand, one postulates a collapse directly at the level of the whole system. So, regardless of the specifics of the initial state, the system will be driven to an eigenstate of the collapse operator. The strength of this collapse is always governed by the parameter λ_{CSL}, again, regardless of the details of the initial state, i.e., of there being or not special correlations in the state.

Moreover, in standard CSL, the collapse rate grows as N does because the collapse operator is constructed as a sum of single particle operators. In our model, this is not the case because Φ of the total system is not the sum of single particle operators. As we said before, this might seem problematic because our model does not appear to be able to distinguish micro and macro states, and to treat them accordingly. In order to deal with this, we can just *postulate* the collapse rate to be a function of the number of particles in the system; in other words, we postulate that the collapse rate has to be *renormalized* when changing the scale of the system under study. This might seem odd, but of course there is no guarantee that the laws of nature are such that the values of their parameters are scale-invariant (and, in fact, this even happens in quantum field theory with the change, or *running*, of the value of the parameters at different scales). Alternatively, instead of renormalizing λ as described above, one could simply add an N to the definition of our collapse operator. In fact, given that Φ is meant to measure consciousness, and thus, expected to capture some sort of complexity within the system, that N might even naturally appear in the definition of Φ.

4.2 Interpreting the Model

The standard interpretation of quantum mechanics subscribes to the so-called Eigenvalue-Eigenvector (EE) rule, which holds that a physical system possesses the value α for a property represented by the operator O if and only if the quantum state assigned to the system is an eigenstate of O with eigenvalue α. Such a rule is essential in order to link the mathematical apparatus of the standard formalism and predictions, i.e., it plays the role of a (partial) physical *interpretation* of the theory.

A legitimate concern regarding collapse models arises from the fact that such formalisms lead systems to states which are very close to eigenstates of the collapse operator, but not exactly to such eigenstates. Therefore, if one continues to subscribe to the EE rule, systems under collapse dynamics never actually possess well-defined values for the property associated with the collapse operator—ostensibly in contrast with what one was hoping for. As a result, one cannot *interpret* collapse theories in terms of EE and a different interpretation—that specifies the relation between the mathematical and the physical objects—is required.

In the context of standard collapse models, one could solve this issue with the use of the, so-called, *fuzzy link* interpretation introduced in Albert and Loewer (1996), in which one allows for some tolerance away from an eigenstate to ascribe the possession of well-defined properties. There are, however, complications with this approach. First off, it is not clear how to define in a non-vague and non-arbitrary way how close a state needs to be to an eigenstate in order for the value of a property to be well-defined. Moreover, this type of interpretation remains as ontologically obscure as the standard interpretation because it only talks about possession of properties but remains silent regarding what are supposed to be the property bearers according to the theory.

A more atractive alternative, again, in the context of standard collapse models, is to construct out of the wave function a, so-called, primitive ontology and to interpret it as the stuff that populates the world (see Allori (2005)). The most popular options in this direction are the flash ontology, in which the centers of the GRW collapses are taken to constitute the primitive ontology, and the mass density ontology, available both for GRW and CSL, in which a mass density in 3D space is constructed out of the wave function as the expectation value of the mass density operator. While promising, these approaches face some open issues (see, e.g., McQueen (2015)).

As with standard collapse models, one may worry about the fact that the SOC model does not really lead systems to eigenstates of Φ, but only to states that are very close to those eigenstates. Also, as with standard collapse models, if one adheres to the EE rule, one gets into trouble because one concludes that SOC leads to a scenario in which conscious states never actually occur. The solution, as with standard collapse models, is to deviate from the EE rule and introduce some alternative. Following what we said above, one option would be to introduce some sort of fuzzy link that ascribes consciousness to states which are close enough to Φ eigenstates. However, as with the standard case, it seems difficult to rigorously define such a "close enough." The other option, of course, is to introduce a primitive ontology, such as mass density, and use it to interpret the SOC model.

There are of course a number of concerns with collapse models, such as the so-called tails problem, that are not specific to our model. In fact, any proposal to solve the measurement problem on the market nowadays contains at least some open issues. We would be satisfied if our model turns out to be no worse than standard collapse models.

Acknowledgments

EO acknowledges partial financial support from DGAPA-UNAM project IN102219.

References

Adler, S. L. (2007). Lower and upper bounds on CSL parameters from latent image formation and igm heating. Journal of Physics A: Mathematical and Theoretical, 40(12):2935.

Albert, D. Z. and Loewer, B. (1996). Tails of schrodinger's cat. In Clifton, R., editor, Perspectives on Quantum Reality: non-relativistic, relativistic, field-theoretic, page 81–92. Kluwer.

Allori, V. (2015). Primitive ontology in a nutshell. International Journal of Quantum Foundations, 1:107–122.

Averill, E. W. and Keating, B. (1981). Does interactionism violate a law of classical physics? *Mind*, 90:102–107.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

Bassi, A. and Ghirardi, G. C. (2003). Dynamical reduction models. Phys. Rep., 379:257–426.

Bassi, A., Lochan, K., Satin, S., Singh, T. P., and Ulbricht, H. (2013). Models of wave-function collapse, underlying theories, and experimental tests. Rev. Mod. Phys., 85:471.

Bell, J. S. (1990). Against measurement. In Miller, A. I., editor, Sixty-two Years of Uncertainty. Plenum Press.

Block, N. (2002). Some concepts of consciousness. In Chalmers, D., editor, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.

Block, N. (2011). The higher order approach to consciousness is defunct. *Analysis*, 71(3):419–431.

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences*, 18(9):445–447.

Chalmers, D. (2010). *The Character of Consciousness*. Oxford University Press.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D. and McQueen, K. J. (MS). Wave-function collapse theories of consciousness. https://www.youtube.com/watch?v=UL1h-QgeD9c, https://www.youtube.com/watch?v=R-jOfW9UIEA\& feature=youtu.be\&list=PLl_UXfN1hubVda8RyXj1FLgwK4tW_AqEA\&t=2016.

Dagasperis, A., Fonda, L., and Ghirardi, G. (1974). Does the lifetime of an unstable system depend on the measuring apparatus? *Il nuovo cimento A.*, 21(3):471–484.

Dennett, D. C. (1991). *Consciousness Explained*. Back Bay Books.

Everett, H. (1957). 'relative state' formulation of quantum mechanics. Rev. Mod. Phys., 29(3).

Feldmann, W. and Tumulka, R. (2012). Parameter diagrams of the GRW and CSL theories of wavefunction collapse. Journal of Physics A: Mathematical and Theoretical, 45(6):065304.

Ghirardi, G. C., Rimini, A., and Weber, T. (1986). Unified dynamics for microscopic and macroscopic systems. Phys. Rev. D, 34:470–491.

Goldstein, S. (2013). Bohmian mechanics. In Zalta, E. N., editor, The Stanford Encyclopedia of Philosophy

Hutto, D. and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content.* MIT Press.

Larmer, R. (1986). Mind-body interactionism and the conservation of energy. *International Philosophical Quarterly*, 26:277–285.

Lewis, D. (1978). Mad pain and martian pain. In Block, N., editor, *Readings in the Philosophy of Psychology*, volume 1. Harvard University Press.

London, F. and Bauer, E. (1939). La th orie de l'observation en m canique quantique. Actualit s scientifiques et industrielles, 755.

McQueen, K. J. (2015). Four tails problems for dynamical collapse theories. Studies in History and Philosophy of Modern Physics, 49:10–18.

Noë, A. (2005). *Action in perception*. Bradford Books.

Okon, E. and Sebastián, M. Á. (2018). A consciousness-based quantum objective collapse model. *Synthese*.

Pearle, P. (1989). Combining stochastic dynamical state vector reduction with spontaneous localization. Phys. Rev. A, 39:2277–2289.

Pearle, P. (1990). Toward a relativistic theory of statevector reduction. In Sixty-Two Years of Uncertainty, pages 193–214. Plenum Press.

Prinz, J. (2012). *The Conscious Brain*. Oxford University Press.

Sebastián, M. Á. (2014). Dreams: An empirical way to settle the discussion between cognitive and non-cognitive theories of consciousness. *Synthese*, 191(2):263–285.

Sebastián, M. Á. (2015). What panpsychists should reject: On the incompatibility of panpsychism and organizational invariantism. *Philosophical Studies*, 172(7):1833–1846.

Stapp, H. (2007). *Mindful Universe*. Springer.

Wigner, E. (1967). Remarks on the mind-body question. In Symmetries and Reflections, page 171–184. Indiana University Press.