Does consciousness collapse the quantum state?

P. A. Knott*

Few human endeavours have provided more excitement - and more confusion - than the studies of both consciousness and quantum mechanics. Perhaps, then, it is no surprise that the two have, on several occasions, been brought together to attempt to solve core problems regarding the universe in which we live. This article focuses on the so-called measurement problem of quantum mechanics, which lies at the very heart of the theory, and has sparked furious philosophical debate since the conception of quantum mechanics over 100 years ago. The measurement problem can be introduced as follows: The Schrödinger equation predicts that objects in quantum mechanics can have multiple properties simultaneously – we say that these objects are in a superposition state. Many manifestations of this – such as the prediction that an electron can be in two places simultaneously – have now been proved repeatedly in experiments. But the Schrödinger equation can give far more radical predictions: Schrödinger himself proposed a thought experiment in which a cat can be put in a superposition of being dead *and* alive at the same time. And here lies the problem: we clearly do not *see* dead and alive cats roaming our streets, and in the above-mentioned experiment, we only ever measure the electron in one location. To rectify this, the "collapse postulate" has been introduced to quantum theory: On observation a superposition state *collapses*, meaning that only one outcome of an observation, or a measurement, is ever observed. But what exactly constitutes an observation, and what entities can take the role of the observer?

One possible answer to this is that *consciousness* is responsible for collapse in quantum mechanics (Wigner 1961, London & Bauer 1939, Schrödinger 1945, Miranker 2002, Bierman 2003, Chalmers & Mcqueen n.d., Kremnizer & Ranchin 2015, Squires 1996, Stapp 2004, 1996, Germine n.d., Gao 2008, Altaisky 2016, Goswami et al. 1993, Thaheld 2005, Rosenblum & Kuttner 2011). We never *see* a superposition because our consciousness itself *collapses* the superposition. As well as being conceptually attractive, giving consciousness this special role is completely consistent with all experiments in quantum mechanics to date. However, there are numerous other theories of what causes collapse that are equally compatible with our experiments (I will introduce some of these later). And so, until our experiments become more advanced, it seems that the only method to choose between the various interpretations is a philosophical enquiry. With this in mind, the purpose of this article is to ask: *what would happen if we did live in a universe in which consciousness causes collapse*? I will argue – using results from both cognitive

*Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems (CQNE), School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

science and evolution theory – that this assumption leads to numerous fantastical and often absurd predictions. Therefore, while we cannot currently disprove that consciousness causes collapse (CCC), believing in such a theory is accompanied with believing in the radical universe that I outline below. Furthermore, when the CCC thesis is studied in the context of our biological evolution, I present arguments that might cause major problems for the theory, or at least place strong restrictions on how it should be formalised.

The measurement problem and CCC

I will begin by introducing the measurement problem. (For those unfamiliar with quantum mechanics, I introduce the necessary details in Appendix A.) Suppose we have a qubit – a quantum bit that can exist in two states, which we label with the notation $|0\rangle$ and $|1\rangle$. It is now routine in quantum physics labs to prepare such a qubit in a superposition state (normalisation is ignored throughout):

$$|0\rangle + |1\rangle. \tag{1}$$

There are numerous physical systems that can be prepared in this superposition state: a photon in a superposition of two different optical fibres; an electron in a superposition of spin up and spin down; or even the current in a circuit in a superposition of travelling in two different directions. Imagine now we have a device that can accurately measure the state of the qubit. The state of the measuring device before performing the measurement is $|M_{ready}\rangle$. If the qubit is in state $|0\rangle$, then the device displays "0" on its display screen – we say that the state of the measuring device when displaying "0" is $|M_0\rangle$. Likewise, the state $|1\rangle$ is recorded by the measuring device changing into state $|M_1\rangle$. We can then write the interaction between the measuring device and the qubit as:

$$|0\rangle|M_{ready}\rangle \rightarrow |0\rangle|M_0\rangle, \qquad |1\rangle|M_{ready}\rangle \rightarrow |1\rangle|M_1\rangle. \tag{2}$$

Now, the linearity of the Schrödinger equation in quantum mechanics allows us to treat each part of a superposition separately. Therefore, if we measure the superposition state of the qubit in equation (1) the following transformation takes place:

$$(|0\rangle + |1\rangle)|M_{ready}\rangle \rightarrow |0\rangle|M_0\rangle + |1\rangle|M_1\rangle. \tag{3}$$

Now suppose a human observer, initially in the state $|H_{ready}\rangle$, looks at the display screen of the measuring device. We can write down the state of a human observer who has seen 0 on the display screen as $|H_0\rangle$, and similarly for $|H_1\rangle$. Note that these states represent the complete state of the human observer; i.e. the state $|H_{ready}\rangle$ represents the whole body and brain, all the cells and organs, every particle and electrical signal, of a human observer who has seen 0 on the display screen. Of course we could not actually calculate exactly what this state is, but for our purposes it is sufficient to write $|H_{ready}\rangle$. Following the same argument as above, if the qubit is in state $|0\rangle$ then we find

$$|0\rangle|M_{ready}\rangle|H_{ready}\rangle \rightarrow |0\rangle|M_0\rangle|H_0\rangle; \tag{4}$$

the state $|1\rangle$ follows as expected. If the qubit is in a superposition, then the following transformation takes place:

$$\begin{aligned}&(|0\rangle|M_0\rangle + |1\rangle|M_1\rangle)|H_{ready}\rangle \\ &\to |0\rangle|M_0\rangle|H_0\rangle + |1\rangle|M_1\rangle|H_1\rangle.\end{aligned} \tag{5}$$

This state represents a superposition of the observer seeing 0 (whilst the qubit is 0 and the measuring device displays 0) **and** the observer seeing 1 (whilst the qubit is 1 and the measuring device displays 1). The human observer is therefore in a superposition of two different states.

Before discussing this further, it is worth mentioning that whilst this experiment might seem out of reach to most readers, in fact you can perform this experiment right now. Just search the web for "quantum random number generators", and numerous websites will give you the chance to randomly generate a number 0 or 1 (try to find a site that generates these numbers live). These quantum random number generators work the same as the experiment above: a quantum state is prepared in a superposition of two states, then this state is measured. In this case, the state $|M_0\rangle$ will represent the complete state of the measuring device, as well as your computer monitor displaying the result 0, and everything relevant in between, such as the information travelling over the Internet. The same result will hold by following the arguments above, and the final state will be given by equation (5). Therefore, according to the calculation above, just by running one of these quantum random number generators online, **you** will be put into a superposition of seeing 0 and 1!

Alarm bells might be ringing now: our intuition very clearly tells us that a human cannot be put into a superposition, especially not this easily. But if we open most undergraduate quantum mechanics books, a solution to this can be found: One postulate of quantum mechanics state that *only isolated systems* evolve with the Schrödinger equation. There is then an additional postulate: when a quantum state is *measured*, it collapses into just one of the states in the superposition. But this "collapse postulate" raises as many questions as it answers: What constitutes a measurement? How does the collapse happen? And if measuring devices and humans are made of quantum particles (atoms etc), then why should they behave differently from the things they measure? These questions together constitute *the measurement problem*.

Numerous attempts have now been made to transform the collapse postulate from an unexplained statement to a more complete theory, with the goal of solving the measurement problem. Theories now exist in which complexity causes collapse (Ghirardi et al. 1986); or gravity causes collapse (Penrose 1994); or, most importantly for this essay, *consciousness causes collapse* (CCC). There are different possibilities for exactly *how* consciousness causes collapse; here I take a minimalist approach and provide two axioms – this will encompass most varieties of CCC. The axioms are: i) Systems isolated from any consciousness will evolve with the Schrödinger equation, and ii) in the presence of consciousness, a superposition state collapses into just one of the constituent states of the superposition. Equations (3) and (5) demonstrate how these axioms operate: assuming a measuring device or qubit is not conscious, equation (3) contains no consciousness, and therefore as the Schrödinger equation governs this evolution (according to postulate i), equation (3) still holds. On the other hand, a conscious entity – in particular a conscious human observer – is present in the interaction in equation (5). Therefore, collapse will occur:

$$\begin{aligned} &|0\rangle|M_0\rangle|H_0\rangle + |1\rangle|M_1\rangle|H_1\rangle \\ &\rightarrow \quad |0\rangle|M_0\rangle|H_0\rangle \quad or \quad |1\rangle|M_1\rangle|H_1\rangle. \end{aligned} \tag{6}$$

The outcome is that you either observe 0, or 1, but not both. This already raises some interesting questions: Is the collapse instantaneous, or does consciousness make the state unstable, so that it has a certain probability of collapsing in a given time? And precisely how is consciousness defined? For the latter question, my main conclusions in this essay will hold true for all definitions of consciousness known to me, except those definitions that advocate panpsychism (in panpsychism everything is conscious, to different degrees, generally with humans being the "most" conscious entities), though I discuss later that similar conclusions hold with panpsychism. To follow the arguments more easily, it might be beneficial to keep a definition of consciousness in your head that implies that all "advanced" animals are conscious, but it is up to you where you draw the line. For the former question, whether collapse is instantaneous or probabilistic does not affect my main conclusions; but for ease of explanation (and for dramatic effect) I will adopt the notion that consciousness causes instantaneous collapse.

If CCC is correct then this leads to the comforting conclusion that humans are never in superposition states. Despite this, by looking at equation (3) we can already draw some less comforting conclusions: this predicts that a measuring device can be in a superposition state. Therefore, if you carry out a quantum random number generator experiment, then *before you look* at your computer monitor it will be in a superposition of displaying 0 and 1. Any number of simple thought experiments can illustrate just how bizarre this conclusion is. For example, imagine you create a simple computer program that allows your computer to automatically run this online quantum random number generator experiment, and then print the result, 0 or 1, onto a piece of paper. Now set your computer going, and leave the room *before* the experiment result is displayed on the screen or printed (e.g. introduce a delay between starting the program and generating the number). Assuming there is nothing conscious in the room, the end result will be a physical piece of paper that is in a superposition of being imprinted with ink in two different shapes, 0 and 1. But then, as soon as you enter the room, *the quantum state collapses* leaving a single piece of paper displaying 0 or 1. We can let our imagination take us further still: imagine connecting your computer to a drone programmed to fly north if the result 0 is obtained, and south if the result 1 is obtained. Before anything conscious sees the drone, it will be flying in *both* directions simultaneously!

Putting our unconscious into a superposition

Things get even more bizarre when this same analysis is applied to experiments in visual psychology and cognitive science. It is now possible for experiments to make things disappear in front of our eyes. Such an experiment described in (Koch & Hepp 2006) involves placing a stationary image in front of one eye of a subject, whilst a rapidly flickering image is placed in front of the other eye. With a carefully tuned rate of flickering, the stationary object becomes invisible to the subject, who then only becomes consciously aware of the flickering image.

Koch and Hepp (Koch & Hepp 2006) then proposed how this trick can be used to probe the connection between consciousness and quantum mechanics. They suggested to put the stationary image in a superposition state, e.g. using the computer monitor introduced above that is in a superposition of displaying 0 and 1. The flickering is then tuned so that the stationary image doesn't enter into the consciousness of the observer. Despite this, the light from the image *is still entering into the eye* of the observer, and following the same arguments as in equation (3), where in this case the eye is the measuring device, the rods and cones in the eye will be put into a superposition of collecting the light from the screen that displays 0, *and* collecting the light from the screen that displays 1. The eye itself is not conscious, and according to CCC this means that collapse cannot occur here. So whilst our consciousness cannot be in a superposition, CCC predicts that our eyes can. In fact, this will not stop at the eyes: a signal will still travel from the eyes into the brain, and this signal will be a superposition of encoding 0 and 1. Only when the signal reaches our consciousness will the superposition state collapse, at which point we are consciously aware of seeing 0 or 1, not both (see Appendix B for a more detailed description of this).

This experiment will be challenging with current technology, but in the near future it may well be possible. There are now a number of proposals for how light prepared in a superposition of two different states could actually be detected by the human eye, thus enabling a masking experiment similar to that introduced above (Rahnama et al. 2009, Phan et al. 2014, Dodel et al. 2016, Sekatski et al. 2009, Vivoli et al. 2016).

The thought experiment introduced above shows that in CCC, while our consciousness cannot be in a superposition, our eyes can, and a signal entering our brain can also be in a superposition state. But how far into the brain can such a superposition state travel? We are already aware from our everyday experience that our brain processes many things unconsciously; for example, are you consciously aware of the movement of your toes as you read this? There is now a wealth of evidence from a variety of fascinating experiments in psychology and brain imaging, many of which are summarised in (Dehaene 2014), that have now shown that unconscious/subliminal processing reaches far into the brain. In fact, "virtually all the brains regions can participate in both conscious and unconscious thought" (Dehaene 2014). This has been shown with indirect evidence, as well as direct brain scans and even electrodes inserted directly into the brain. It is now known that, without us being consciously aware, we can do maths, evaluate chess boards, and process the meaning of words and phrases. A particularly illuminating case study comes from blindsight patients, who cannot consciously see and will deny that any visual information is entering their mind, yet can still point to a light or navigate through a room. All of these experiments involve information entering the brain, normally through the eyes. It is therefore in principle possible that this visual input be replaced with input in a quantum superposition, thereby putting our unconscious into a superposition state, at least according to CCC. It has now even been shown that we can be unconsciously sexually aroused (Oei et al. 2012, Wernicke et al. 2017, Jiang et al. 2006); CCC therefore predicts that we can unknowingly be in a superposition of arousal states!

Koch and Hepp have suggested that such experiments can be used as a test of CCC, but by looking in more detail about how such a test would be implemented, we can see that this claim is false. In (Wernicke et al. 2017), direct brain scans confirm that our unconscious can be accessed without us being consciously aware of this. If our unconscious is in a superposition state, then the results of the brain scan will also be in a superposition. So far so good: the outcome of the experiment encodes that there is a superposition. But as soon as a conscious observer tries to confirm that there is a superposition by looking at the results of the experiment, according to CCC collapse will occur. Whether CCC is correct or not, the final result of such an experiment will be the same: the final observer will only ever see one outcome. This doesn't mean that CCC cannot be tested, but as I discuss in (Knott 2018) the only way to test whether a macroscopic state, such as a brain, is in a superposition is to perform an interference experiment. And while it is now routine to demonstrate interference of atoms, and cutting-edge experiments can even interfere C60 molecules (Arndt et al. 1999), it is currently beyond comprehension to perform a quantum interference experiment on something as complex and dynamic as a human brain. The only hope for this may be to simulate a human brain on a quantum computer, but this is still in the realm of science fiction. (See Appendix A for more on this.)

Naturally occurring superpositions

Up to now I have only discussed superposition states that are formed artificially in experiments, but in fact superpositions are created almost continuously in our surroundings. For example, when orbiting a nucleus an electron exists in a perpetual superposition state. In some sense it does have a precise energy, but the actual position of the electron is indefinite. A free electron – one not trapped in an atom or by some other potential well – has a wavefunction that extends over all space. In the terminology of superpositions, such an electron is in a superposition of being everywhere in the universe (except those positions which would require an infinite energy to enter into). Another example is that of radioactive decay. It we take one single atom of a radioactive isotope, and place it next to a Geiger counter, then it is normally said that there will be a certain probability that in a given time the atom will decay and the counter will click. For example in 1 minute there is a 50% chance of an oxygen-14 atom decaying. However, this description in terms of probabilities doesn't give the full picture. This process is a random quantum event, and in fact the atom decays and does not decay in a superposition (the Schrödinger's cat thought experiment gives an interesting application of this). CCC would then predict that the Geiger counter would click and not click in superposition. Only when the signal reaches our consciousness would the state collapse. Furthermore, because there is a non-zero quantum probability of the atom decaying at any given time, the atom is actually in a superposition of decaying at any given time – it is a continuous superposition of different possibilities. Macroscopic objects like the Geiger counter may not actually enter into a continuous superposition – this is a subtle and complex issue beyond the scope of this essay (Wallace 2012). But for our purposes it is sufficient to say that in this example the Geiger counter will enter into a superposition of an unimaginably large number of different states, each corresponding to the Geiger counter clicking at a certain time.[1]

Of course it is not surprising, and nor is it contested, that quantum-scale objects can be in superposition states. But what about macroscopic objects – are there situations in which macroscopic superpositions are created naturally? One such mechanism is when a macroscopic phenomena depends on a critical point. For example, a black hole is formed when the mass density of a collection of matter, such as a collapsing star, reaches a certain threshold (Choptuik 1993). A collection of matter on the periphery of becoming a black hole can, at some given instant in time, be pushed over the critical density by the mass of a single electron. Therefore, a passing electron in a superposition state (of being close to and far away from the collection of matter) can force the black hole to form, and not form, in superposition. This may seem like an unlikely event, but as introduced above quantum-scale objects generally exist in superposed states, and therefore events that rely on overcoming a critical barrier will invariably happen, or not, in superposition. Chaotic systems are another illuminating example. Here, a small change in initial conditions causes a large change in the final outcome. If the initial conditions rely on the properties of quantum-scale objects, then such chaotic systems provide a mechanism of creating macroscopic superpositions.

Evolution can also happen in superposition. It can take just a single high-energy photon to cause a mutation in DNA (Northrop 2001); and photons are produced in quantum-random events such as radioactive decay and therefore often exist in superposition states. A photon in a superposition of two different locations can then cause a mutation, and not cause a mutation, in superposition. And when such a mutation, or a collection of mutations, is responsible for a critical divergence of a species, then this too can happen in a superposition.

These examples were intended to give intuitively plausible mechanisms whereby a quantum-scale superposition can naturally become a macroscopic superposition. But in fact, if there is nothing available to collapse the quantum state, then macroscopic superpositions will be abundant throughout the universe. To see this, imagine the early universe: a high-energy soup of quantum mechanical particles. As mentioned above, the wavefunction of a free particle quickly evolves into a superposition state, and this implies that particles in the high-energy quantum soup will generally all exists in superposition states. It seems reasonable to suppose that changes in the early state of the universe will lead to radically different present-day universe states. And as the early state can be in a superposition, so can the final state.

Putting all this together paints a radical and quite mind-boggling picture of the universe. If consciousness is responsible for collapsing the wavefunction, then before

[1] Quantum random processes govern not only radioactive decay, but also nuclear reactions, chemical reactions, electromagnetic interactions, and so on.

any conscious entities evolved, macroscopic superpositions are likely to have been everywhere. Imagine a solar system in which each planet does not follow a single trajectory, but is in a superposition of different trajectories. Not only planets, but asteroids, stars, and even galaxies will have existed in multiple states simultaneously. On one such planets, on one branch of the vast macroscopic superposition state, the conditions would have been right for simple organisms to begin their evolutionary journey towards conscious beings. Then, due to the bombardment of photons originating from the distant quantum events, mutations would drive the evolution of these organisms towards growing complexity. But these mutations would happen in superposition, and therefore these organisms would be in a superposition of different evolutionary states. We can imagine flowers in superposition of different colours, spiders in superposition of different numbers of legs, and dinosaurs in superposition with different shaped horns. Of course organisms with unsuitable mutations would still die out. This explanation all depends heavily on where you draw the line for consciousness, but we could even imagine a superposition-branch where dinosaurs didn't die out, and a near infinite number of worlds with creatures beyond our imagination. And then, after all this, the first conscious creature evolves and enters into the world and suddenly, 'POP', this great macroscopic superposition state collapses, leaving no trace of what came before.

Potential problems with CCC

This description certainly does not disprove a theory of CCC, but if we continue our imaginative exploration we can uncover some potentially fatal flaws in the theory, or at least significant restrictions it must abide to. To see this, we can ask: under what conditions does consciousness cause collapse? There seem to be two alternatives here: i) a single conscious entity cannot enter into a superposition, and therefore in circumstances in which a conscious entity would otherwise enter into a superposition, the wavefunction collapses, leaving a single conscious entity; or ii) when any superposition state interacts with a conscious entity, the superposition collapses.

To explore the first option, imagine our universe before any consciousness evolved. Picture a world almost identical to ours, except that one distant star has been shifted slightly so that, even though the star's position would be unnoticeable to us, the light reaching the Earth would in principle distinguish this world from ours. Now take a superposition of our world with that world. We have seen above that all kinds of macroscopic superpositions will be abundant, and so it seems completely plausible – maybe even inevitable – that this superposition state would arise. If this thought experiment is set up carefully enough, the evolutionary chain towards consciousness could be near identical. What then would happen when the first conscious entity comes into existence? Here we have a situation where there is a superposition of two conscious minds – one on each of the superimposed worlds. But it is not the case that one consciousness has entered into a superposition, so if i) is correct then collapse will not happen. If we fast forward evolution to humanity, then this will lead to all humans being in a superposition, including you and I. We then have a theory in which consciousness causes collapse, but we can still have a superposition of conscious minds!

This goes against the central motivation of proposing CCC, namely that it seems implausible that a consciousness can be in a superposition, and therefore consciousness must cause the collapse. Mechanism ii) would alleviate this problem because, in this case, as soon as the first conscious entity comes into existence collapse would happen. Thus a superposition of different evolutionary branches containing conscious entities would be prevented. But this mechanism brings its own difficulties, as illustrated in the following example. Double slit experiments with streams of single particles have now proved many times that a single particle can enter into a superposition and pass through 2 slits simultaneously (see Appendix A). The usual explanation of this is that the wavefunction of the particle extends far enough to cover the spatial region of both slits. But this stops short of the full description of the particle: solving the Schrödinger equation will demonstrate that the wavefunction of the particle in fact extends over *all* space (except regions that form an infinite potential barrier, which are unphysical anyway). But all space includes the space within the brain of the person performing the experiment; the particle is in a superposition of being in the slits and in the brain, at the same time. Phenomena of this sort are rarely discussed because the probability of detecting the particle inside your brain is so close to 0 that it would likely take longer than the age of the universe till an experiment could actually confirm this. Now if the particle is an electron, then the electron can be in a superposition of being inside the slits *and* somewhere in your brain that forms part of your conscious processing, for example in an electrical synapse between relevant neurons. But if ii) is the correct mechanism, then this electron has interacted with consciousness, and therefore collapse will occur. With a probability of practically 1, the electron will collapse to being inside the slits rather than in the brain, but unless there is some elaborate mechanism that only causes partial collapse, the position wavefunction should collapse completely. The electron is therefore in just one of the slits. But this goes against the experimental data itself – we know that the electron can pass through both slits because it interferes with itself. A successful theory of CCC would have to overcome issues such as this, for example by postulating that the property of collapse depends on the amplitude of the part of the wavefunction that interacts with consciousness.

For CCC to be successful a number of other problems must be addressed: How does the CCC mechanism know which basis to collapse into? Why would something that according to CCC plays such a fundamental role in physics (consciousness) evolve naturally by a process (Darwinism) that seemingly has nothing directly to do with the physical laws themselves? And, as discussed recently by McQueen and Chalmers (Chalmers & Mcqueen n.d.), how can consciousness evolve if it continuously causes collapse?

I will now briefly discuss how a panpsychist theory of consciousness (in which everything is conscious to differing degrees) would operate within CCC, for example integrated information theory causing collapse (Kremnizer & Ranchin 2015). Such a theory could follow a principal like: *the more conscious an entity is, the more likely it is to collapse the wavefunction*. Experiments have proved that C60 molecules

(Arndt et al. 1999), and it has recently been announced that up to 50 qubits, can be put into a superposition in a quantum computer (*The future is quantum - IBM Blog Research* n.d.). Therefore, these systems are clearly not conscious enough to cause collapse. Quantum computers will likely provide the strictest restrictions to such theories for now – arguably error correction codes have a large integrated information (Aaronson n.d.) – and at the current impressive rate of technological progress increasingly complex systems will be ruled out as not conscious enough to cause collapse. With this in mind, all of the discussions in this essay could be revisited, raising the following questions: Is your desktop computer conscious enough to cause collapse? What about your eyes? Or your subconscious? Or simple but macroscopic objects? Or creatures early on the evolutionary chain? But in a sense it doesn't matter *how* conscious these objects are – they cannot cause collapse 100% of the time because otherwise everything would cause collapse. Therefore, even in these theories there will be a nonzero probability that your computer, your subconscious, and even your conscious mind itself, can be in a superposition. Even if such a probability is practically zero, would such a conclusion be satisfactory?

Discussion

At first sight, the theory that consciousness causes collapse seems to provide an attractive and intuitive resolution to the measurement problem: *I know from my experience that my conscious mind does not enter into a superposition, but this is resolved if consciousness causes the superposition to collapse*. However, we have seen here that the universe that is predicted by CCC is far from intuitive. Fortunately, there are a wide spectrum of alternatives to CCC. If we still wish to believe that *something* collapses the wavefunction, then there are theories that gravity (Penrose 1994) or complexity (Ghirardi et al. 1986) are responsible for this. But all these collapse theories have the attribute that they modify the highly regarded and never disproved Schrödinger equation. This is unsatisfactory for many, and it is now more common among quantum physicists to believe that the Schrödinger equation is sufficient as it is. In particular, as I discuss elsewhere (Knott 2018), the well-established framework of decoherence (Zurek 2003, Schlosshauer 2005) can now explain why the Schrödinger equation alone can lead to the appearance of collapse, and can therefore explain why macroscopic objects are never seen in superposition states.

But if nothing collapses the wavefunction, then what kind of reality do we live in? The answer to this depends in a large part on how much of a realist you are. At one end of the spectrum, if you believe that the whole quantum state is "real" then this leads to the Everett interpretation, a.k.a. many worlds (Everett 1957, Wallace 2012, Tegmark 2014). The other extreme is to believe that quantum mechanics is just a tool for predicting outcomes to experiments and can tell us nothing about reality, e.g. QBism (Fuchs & Stacey 2016, Mermin 2014, Caves et al. 2002). Distinguishing between these different alternatives is at present a purely philosophical issue – while future experiments can in principle determine between collapse and non-collapse models (see Appendix A), it is not clear whether experiments could ever distinguish between the other alternatives. Attributes such as plausibility then become important, and advocates of CCC must stomach a pretty bizarre picture of our universe, in which some regions of our brain can be in a superposition but others cannot; where entire species exist and don't exist simultaneously; and perhaps most importantly, a universe in which our consciousness is endowed with the power to instantaneously collapse brains, computers, entire species, and even stars and galaxies!

Acknowledgements: The author would like to thank Loz Leigh, Adam Barrett, Andrew Knott, Lewis Clark, Konstantinos Meichanetzidis, Phil Knott, and Jesús Rubio for inspiring discussions and helpful suggestion. This work was supported by the Foundational Questions Institute (fqxi.org) under the Physics of the Observer Programme (Grant No. FQXi-RFP-1601).

References

Aaronson, S. (n.d.), 'Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander)'.
URL: *https://www.scottaaronson.com/blog/?p=1799*

Altaisky, M. V. (2016), 'Consciousness and quantum mechanics of macroscopic systems', *arXiv preprint arXiv:1607.04400* .

Arndt, M., Nairz, O., Vos-Andreae, J., Keller, C., van der Zouw, G. & Zeilinger, A. (1999), 'Waveparticle duality of C60 molecules', *Nature* **401**(6754), 680–682.

Bierman, D. (2003), 'Does consciousness collapse the wave-packet?', *Mind and Matter* **1**(1), 45—-57.

Caves, C. M., Fuchs, C. A. & Schack, R. (2002), 'Unknown quantum states: The quantum de Finetti representation', *Journal of Mathematical Physics* **43**(9), 4537–4559.

Chalmers, D. & Mcqueen, K. (n.d.), 'Consciousness and the Collapse of the Wave Function'.
URL: *http://consc.net/slides/collapse.pdf*

Choptuik, M. W. (1993), 'Universality and scaling in gravitational collapse of a massless scalar field', *Physical Review Letters* **70**(1), 9–12.

Dehaene, S. (2014), *Consciousness and the brain: Deciphering how the brain codes our thoughts*, Penguin.

Dodel, A., Mayinda, A., Oudot, E., Martin, A., Sekatski, P., Bancal, J.-D. & Sangouard, N. (2016), 'Proposal for witnessing non-classical light with the human eye', *arXiv preprint arXiv:1611.08542* .

Everett, H. (1957), '"Relative State" Formulation of Quantum Mechanics', *Reviews of Modern Physics* **29**(3), 454–462.

Fuchs, C. A. & Stacey, B. C. (2016), 'Qbist quantum mechanics: Quantum theory as a hero's handbook', *arXiv preprint arXiv:1612.07308* .

Gao, S. (2008), 'A Quantum Theory of Consciousness', *Minds and Machines* **18**(1), 39–52.

Germine, M. (n.d.), 'A Framework for Quantum Paradigms in Psychopathology Based on the One Mind Model of Quantum Reality'.
URL: *https://www.academicpublishingplatforms.com/search-article.php?journal=DPs&number=1&article=1531*

Ghirardi, G. C., Rimini, A. & Weber, T. (1986), 'Unified dynamics for microscopic and macroscopic systems', *Physical Review D* **34**(2), 470–491.

Goswami, A., Reed, R. E. & Goswami, M. (1993), *The self-aware universe : how consciousness creates the material world*, Putnam's Sons.

Jiang, Y., Costello, P., Fang, F., Huang, M. & He, S. (2006), 'A gender- and sexual orientation-dependent spatial attentional effect of invisible images.', *Proceedings of the National Academy of Sciences of the United States of America* **103**(45), 17048–52.

Knott, P. A. (2018), 'Decoherence, quantum Darwinism, and the emergence of an objective reality', *In preparation* .

Koch, C. & Hepp, K. (2006), 'Quantum mechanics in the brain', *Nature* **440**(7084), 611–611.

Kremnizer, K. & Ranchin, A. (2015), 'Integrated information-induced quantum collapse', *Foundations of Physics* **45**(8), 889–899.

London, F. & Bauer, E. (1939), *La théorie de l'observation en mécanique quantique*.

Mermin, N. D. (2014), 'Physics: QBism puts the scientist back into science', *Nature* **507**(7493), 421–423.

Miranker, W. (2002), 'A quantum state model of consciousness', *Journal of Consciousness Studies* **9**(3), 3–14.

Northrop, R. B. (2001), *Noninvasive instrumentation and measurement in medical diagnosis*, CRC press.

Oei, N. Y. L., Rombouts, S. A., Soeter, R. P., van Gerven, J. M. & Both, S. (2012), 'Dopamine Modulates Reward System Activity During Subconscious Processing of Sexual Stimuli', *Neuropsychopharmacology* **37**(7), 1729–1737.

Penrose, R. (1994), *Shadows of the mind : a search for the missing science of consciousness*, Oxford University Press.

Phan, N. M., Cheng, M. F., Bessarab, D. A. & Krivitsky, L. A. (2014), 'Interaction of Fixed Number of Photons with Retinal Rod Cells', *Physical Review Letters* **112**(21), 213601.

Rahnama, M., Salari, V. & Tuszynski, J. A. (2009), 'How Can the Visual Quantum Information be Transferred to The Brain Intact, Collapsing There and Causing Consciousness?', *NeuroQuantology* **7**(4).

Rosenblum, B. & Kuttner, F. (2011), *Quantum enigma : physics encounters consciousness*, Oxford University Press.

Schlosshauer, M. (2005), 'Decoherence, the measurement problem, and interpretations of quantum mechanics', *Reviews of Modern Physics* **76**, 1267.

Schrödinger, E. (1945), *What is life? The Physical Aspect of the Living Cell*, Cambridge: The Univ. Press.

Sekatski, P., Brunner, N., Branciard, C., Gisin, N. & Simon, C. (2009), 'Towards Quantum Experiments with Human Eyes as Detectors Based on Cloning via Stimulated Emission', *Physical Review Letters* **103**(11), 113601.

Squires, E. J. (1996), 'What are quantum theorists doing at a conference on consciousness?', *arXiv preprint quant-ph/9602006* .

Stapp, H. P. (1996), 'Chance, choice, and consciousness: A causal quantum theory of the mind/brain', *Unpublished manuscript* .

Stapp, H. P. (2004), Mind, matter, and quantum mechanics, *in* 'Mind, matter and quantum mechanics', Springer, pp. 81–118.

Tegmark, M. (2014), *Our mathematical universe : my quest for the ultimate nature of reality*, Vintage.

Thaheld, F. H. (2005), 'Does consciousness really collapse the wave function?', *Biosystems* **81**(2), 113–124.

The future is quantum - IBM Blog Research (n.d.).
URL: *https://www.ibm.com/blogs/research/2017/11/the-future-is-quantum/*

Vivoli, V. C., Sekatski, P. & Sangouard, N. (2016), 'What does it take to detect entanglement with the human eye?', *Optica* **3**(5), 473–476.

Wallace, D. (2012), *The emergent multiverse : quantum theory according to the Everett Interpretation*, Oxford University Press.

Wernicke, M., Hofter, C., Jordan, K., Fromberger, P., Dechent, P. & Müller, J. L. (2017), 'Neural correlates of subliminally presented visual sexual stimuli', *Consciousness and cognition* **49**, 35–52.

Wigner, E. P. (1961), 'Remarks on the Mind Body question, in The Scientist Speculates,', *Heinmann, London* .

Zurek, W. H. (2003), 'Decoherence, einselection, and the quantum origins of the classical', *Reviews of Modern Physics* **75**, 715.

A Appendix: Introduction to quantum states, superposition, and interference experiments

Quantum states - Any physical object that is described using quantum mechanics can be expressed as a *quantum state*, using the notation $|\cdot\rangle$. The quantum state for an object contains all the relevant information about that object, and this notation can be used regardless of what the object is. As an example, we can write down the state of a single photon as $|1\rangle$. We can even write down the state of a human using this notation, for example by writing $|H\rangle$. This state contains all the relevant information about this human, including the exact arrangement of all the particles and molecules contained within it - of course we can't realistically know this information, but this doesn't stop us representing the human as $|H\rangle$.

Superposition states - A classical bit can be in one of two states, 0 or 1. We write the equivalent quantum bit, a qubit, as $|0\rangle$ or $|1\rangle$. But one important difference between classical and quantum physics is that a quantum bit can be prepared in a strange state, known as a superposition state, in which it is both $|0\rangle$ and $|1\rangle$ simultaneously. There is no intuitive way of understanding what this state means because there is no analogue in our everyday world, so on some level one has to just accept that a qubit can be both $|0\rangle$ and $|1\rangle$ simultaneously! We write down such a superposition state as

$$|0\rangle + |1\rangle. \tag{7}$$

The $+$ sign can be read as "is in superposition with".

If we have two objects we can put their states together. For example, $|0\rangle|H_0\rangle$ represents a qubit and a human such that the qubit is in the 0 state whilst the human is in the H_0 state.

Interference - How can we prove that quantum mechanical objects can be put into a superposition? The answer lies in the phenomenon of *interference*. Suppose you throw two stones into a pond that simultaneously land close to one another. The waves from the two stones will spread out, and eventually interfere with one another. Every time a peak of the wave coming from one stone meets a peak of a wave coming from the other stone, these peaks reinforce each other - this is constructive interference. Whereas if a peak and a trough meet, they cancel each other out - this is destructive interference.

A similar effect happens when we send a wave through a "double slit", which in effect is a barrier pieced by two small openings (the slits) that are close to one another (search the web for "double slit experiment" for some helpful pictures). Diffraction causes the wave to spread out as it passes through each slit, and this means that the waves coming from the two slits will interfere with one another. Again we get constructive and destructive interference, and this produces an interference pattern that can be measured.

One of the obscure things about quantum physics is that particles can act as waves. This can be seen by sending a stream of electrons through a double slit: an interference pattern is observed, proving that the electrons can have wave-like behaviour. Even more surprisingly, if we send just one electron at a time through the slits, we still get the interference pattern. The reason for this is that the electron passes through both slits simultaneously - i.e. it enters into a superposition of passing through both slits at the same time - and then the two "parts" of the superposition interfere with each other, again producing an interference pattern.

Interferometer - The more general name for a device that can be used to demonstrate interference is an *interferometer*. Interferometers can be used to interfere light, atoms, and even large molecules such as the C60 molecule (Arndt et al. 1999). The general principle of the device is that an object is put into a superposition, and then the different parts of the superposition are made to interfere with each other, producing an interference pattern.

Confirming or ruling out different collapse models - The central equation in quantum mechanics, the Schrödinger equation, predicts that it is in principle possible to put any object into a superposition, and produce an interference pattern. In practice this is extremely challenging, and while cutting-edge experiments can do this with C60 molecules, we are a long long way from putting, for example, a living creature such as a human into a superposition. However, not everyone thinks that this would be even in principle possible: as explained in the main text there are theories in which gravity (Penrose 1994), or complexity (Ghirardi et al. 1986), or most importantly for this essay consciousness, causes the superposition state to collapse. This would prevent us from doing an interference experiment with something like a human.

These theories can in principle be tested: if it can be proved that a massive object can be put into a superposition then this rules out the theory that gravity causes collapse (experiments are currently under way to try to prove or disprove this theory). Similarly, if a conscious mind can be put into a superposition then this would rule out CCC, but to take a physical human mind and put it into a superposition for long enough to show interference is almost unthinkably difficult. One possible alternative would be to simulate a conscious mind on a quantum computer (by definition quantum computers demonstrate interference). This is of course not possible yet, but perhaps in the future such an experiment could be performed.

B Appendix: When would collapse actually happen between the unconscious and the conscious?

For clarity of explanation, in the main text I described that unconscious parts of our brain can be put into a superposition, but when the signal reaches our conscious awareness collapse occurs. In fact in this situation the answer to the question of precisely when the collapse occurs is far more complicated than this. Imagine that using one of the experiments described in the main text we put an unconscious part of your brain into a superposition of two different states, which we can label X and Y. Now it is not yet known precisely how an unconscious state is stored in the brain, but for the sake of this discussion we can imagine that state X is stored in a collection of neurons, and likewise for Y. There is therefore a collection of neurons in a superposition of being in X and Y. In a very short timescale (Tegmark 2014) these neurons will become entangled with surrounding particles and molecules. All that needs to happen for this entanglement-transfer to occur is that information about whether the neurons are in state X or Y gets transferred, but this will happen as soon as any collision or interaction occurs. For example, looking at one individual neuron, if the neuron is in a superposition of firing and not firing, then the magnetic field, or lack thereof, surrounding the neuron will contain information about the state of the neuron. Surrounding particles that are sensitive to this magnetic field will then in turn react, or not react, to this magnetic field, and therefore themselves will encode the state of the neuron. These particles in turn will interact with surrounding particles, and the process will continue.

Now it is not clear exactly in what manner the original state of the superposition of X and Y will become entangled with the rest of the brain, but it is likely that after some timescale the whole brain will be entangled with the original state. The final entangled state will almost certainly not just be a superposition of just two states, but every neuron in the brain will in some way be entangled with the original superposition. Assuming that consciousness is encoded in brain states, i.e. in patterns of neurons firing and not firing, then our consciousness will now be in a superposition. But importantly the different conscious states associated with X and Y will not be noticeably different to us. At this point proponents of CCC have to decide whether CCC will occur if our consciousness is in a superposition of two states that are indistinguishable to us, or whether consciously distinct superpositions are necessary. If the former, then collapse will happen before the unconscious signal reaches our consciousness via the usual information transfer processes in the brain.

Discussions of this sort are clearly quite speculative and cannot be fully clarified without a detailed working model of the brain, but it should be at least clear that, firstly, there will be some timeframe in which our unconscious is in a superposition but our consciousness is not, and secondly that the question of precisely when collapse occurs is extremely complicated.