Quantum consciousness, brains, and cognition

Jerome R. Busemeyer

Indiana University, jbusemey@iu.edu

Meijuan Lu

Meng River Health, ludaifu.88@gmail.com

Abstract

Quantum consciousness concerns both the possible role that quantum mechanics has for understanding consciousness as well as the role that consciousness has for interpreting quantum physics. Quantum brain theories hypothesize that quantum physical processes occur within and between the neurons of the brain and have important effects on cognition as well as consciousness. Quantum cognition is a growing new field in cognitive science concerned with the application of the mathematical principles of quantum theory to human judgment and decision-making behavior. What do all these theories have to do with each other? Quantum theories of consciousness have more to say about quantum physics than cognitive psychology and conscious experiences. Quantum brain theories have not been sufficiently "scaled up" to provide clear implications for how quantum physical processes generate more complex cognition. Quantum cognition theories have avoided addressing fundamental issues about consciousness and have remained agnostic with respect to the quantum brain hypothesis. This article will address the problem of connecting these ideas together by connecting quantum cognition to the other two topics.

Quantum theory is about a century old now, and its reach has grown vast and vigorous, spreading across various fields of science – even outside of physics including brains, cognition, and consciousness. Although the application of quantum theory to these these topics has been investigated separately for many years, connections between them are missing, because they have hardly ever been considered together in one and the same work.[1] For the first time, we review all three applications, and then we describe how they can be connected together. We begin with a discussion of the oldest of the three topics, quantum consciousness.

Quantum consciousness

Why should quantum theory and consciousness have anything to do with each other? One could argue that quantum theory is very mysterious and hard to understand, and so is consciousness, therefore maybe they have something in common? Maybe understanding one could help understand the other? This may not be a really good answer, and the following quote from Dehaene (2014, , pp. 98-99) discussing global workspace theory is a better reason

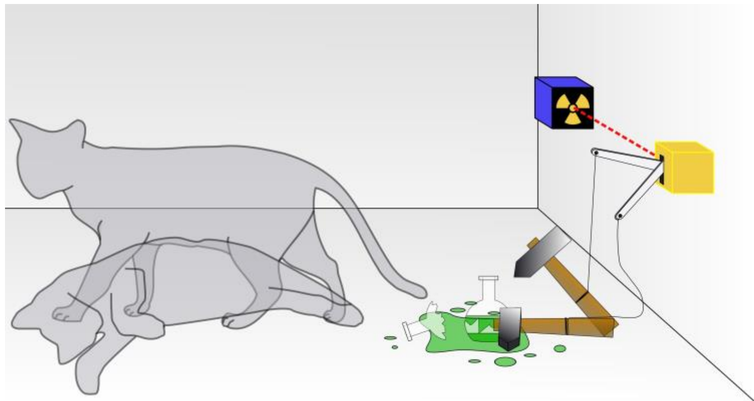
> According to quantum theory, the very act of physical measurement forces the probabilities to collapse into a single discrete measure. In our brain, something similar happens: the very act of consciously attending to an object collapses the probability distribution of its various interpretations and lets us perceive only one of them. Consciousness acts as a discrete measurement device that grants ups a single glimpse of the vast underlying sea of unconscious computations.

Although this quote is relatively recent, the introduction of consciousness into the discussion of quantum mechanics actually occurred much earlier when the its founders encountered the celebrated *measurement problem*.

The measurement problem was illustrated by Schrödinger using his infamous cat experiment, which is a hypothetical experiment involving a cat contained inside the same room as radio active atom and a Geiger counter (see Figure 1). In this situation, there is some chance that the atom decays, and if it does, then it is detected by the geiger counter,

[1]One exception is the encyclopedic review by Atmanspacher (2011).

Figure 1. Schrödinger's Cat Experiment



Dhatfield, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

which then sends a signal to break a flask containing poison. According to quantum theory, the atom starts out superposed between decaying or not decaying. Because the atom is interacting within the same environment as the geiger counter, the flask, and the cat, they all enter into an entangled superposition state with the atom. Consequently, the cat enters an incredulous state that can be described as a superposition of dead and alive. However, we humans never seem to observe this superposition state, and instead, we only observe an event that the cat is definitely dead or definitely alive. Quantum theory has trouble explaining how we come to see just one of the definite states.

Bohr (see Plotnitsky, 2012, for an extensive discussion), along with Heisenberg, Pauli, and others[2], developed a "Copenhagen interpretation" of quantum mechanics, partly to address this measurement problem. Bohr's argument appealed to the limitations of our conscious experiences as human beings. He argued that we may be able to discover mathematical laws of nature that are useful for relating our experiences, but our experiences are limited to the concepts that we have evolved and learned by interacting with the macro world through our human senses. We can only use these macro level concepts (e.g., specifying the orientation of a magnet) to describe the conditions used to measure our observations of the

[2]There are actually several interpretations because Bohr was hard to understand. Also Heisenberg later changed his interpretation to include the idea that potentials become actualized.

micro world (e.g., an electron), and we can only use these macro concepts (record a pointer direction) to report the outcomes of these measurements. This idea that our experiences are limited to macro world concepts does *not* imply that there is one physics for macro world objects, and one physics for micro world objects. There is only one physics and it is quantum all the way from micro to macro. But we can't experience or describe the micro world, and we can only experience and describe what we see and understand in the macro world.[3] In this way, Bohr introduced the importance of our conscious experience into quantum physics.

Von Neumann (1932/1955) addressed the measurement problem by proposing that quantum systems evolve according to two different dynamic operations. One is the continuous evolution of the quantum superposition state according to the Schrödinger differential equation (the type II process), and the other is the probabilistic collapse of the quantum superposition state to a definite outcome that occurs with measurement (the type I process). Von Neumann also postulated what is called the von Neumann chain –all physical systems follow the quantum dynamics, and so the quantum system (e.g., electron) could enter an entangled superposition state with the measurement instrument (e.g., the detector), and the measurement instrument could enter an entangled superposition state with the human brain (e.g., the neurons). The collapse could occur anywhere along this chain, and so the superposition state could persist all the way up through the brain. However, he proposed that our conscious experience lies outside the laws of physics, and so this superposition state must finally stop at the conscious experience. Although he did not necessarily state that consciousness produces the collapse, chain that he postulated suggested that it must stop with consciousness.

Wigner (d'Espagnat, 2005) went a step further by arguing that if the cat in Figure 1 is replaced with his friend (and the poison with something harmless like perfume), so he could later ask his friend whether or not he was superposed while waiting for the atom to decay in the box, his friend would certainly say no – therefore, Wigner concluded, conscious

[3]One reason why we cannot possibly experience a superposition state like the superposition of dead and alive is because the same superposition state can be expressed in an infinite number of different ways depending on the choice of basis vectors.

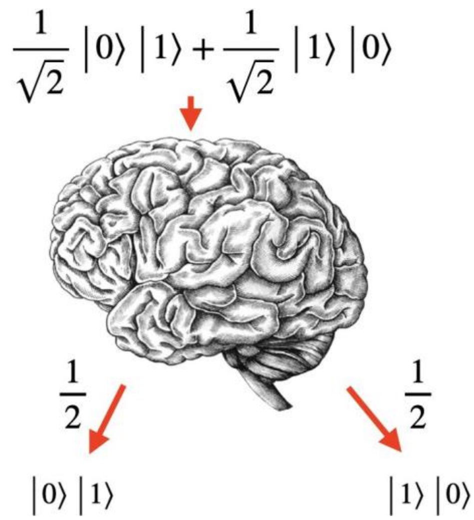
experience (by his friend in the box example) causes the superposition state to collapse.

Later Stapp (1993/2009) added two other types of process, which he called type 0 and III to von Neumann's type I and II processes. Regarding type 0, he pointed out that the observer is a conscious human being who freely chooses what kind of experiment to conduct and what kind of observations to make about the observed system. This free choice of what to observe then changes what is observed according to von Neumann's type I process. The type III process refers to nature's choice of the actualized outcome selected from the superposition of basis states with respect to the measurement basis.[4] Therefore he proposed that a person's conscious choices of what to observe, followed by the conscious experience, finally produce physical neural correlates in the brain.

There are several other more recent quantum consciousness theories (see, e.g., Pylkkannan 2006; Chalmers & McQueen 2022; Gao 2008; Neven et al. 2024), but they all have in common the idea that quantum processes are operating in the brain and consciousness experience is related to superposition and its collapse. Of course, if we demand, as did von Neumann, that all physical processes must be quantum, and brain processes are physical processes, then they must be quantum too. However, most neuroscientists do not see any need for quantum mechanics to understand how the brain works. Instead, they rely mainly on classical physical neural networks to understand perception, cognition, and ultimately consciousness too (see, Seth & Bayne, 2022 for a review of theories of consciousness). Furthermore, most physicists don't agree with the idea that consciousness is needed to produce a collapse of the quantum superposition state. Instead, many of them argue that there is a process called quantum decoherence which produces what appears to an observer as a collapse of the superposition state (see, e.g., Nielsen & Chuang 2000 box 8.4 on p. 387 for a discussion of decoherence with the Schrödinger cat experiment).

[4] Stapp suggested that consciousness could also influence this type III process.

Figure 2. Decoherence problem in the brain. A superposition state (entangled anti correlated Bell state) displayed at the top rapidly decays into a mixture of classical states shown at the bottom.



Quantum Brains

Decoherence is a problem that all quantum brain theories must address (see Figure 2). Decoherence occurs when a quantum state (e.g., an electron) is exposed to a much larger and noisier environmental system (e.g., a magnet measuring its spin). A quantum state that has decohered produces no quantum interference and consequently behaves classically.[5] Quantum computing advantages rely on maintaining a coherent state (that can produce interference effects) for a sufficiently long time to allow the process to complete the required computations. So the problem with the idea that the brain is some type of quantum computer is that, in order to achieve the desired computations, it must avoid rapid decoherence. The brain would need to maintain quantum coherence for sufficiently long time scales (e.g, milliseconds) to be able to perform meaningful quantum computations capable of producing high level cognition. Maintaining quantum coherence normally requires the system to be in a cold, closed, and isolated environment (the way quantum computers are currently built). In contrast, the brain operates in a warm, wet, and noisy environment. Perhaps, however, evolution has found ways to overcome this problem?

[5]Technically, the decohered state produces a probabilistic mixture of classical states.

Despite skepticism, several quantum brain theories have been put forward over the past 25 years. One of the earliest was by Beck & Eccles (1992), who argued that quantum molecular processes, producing exocytosis, occur within the synaptic junctions between neurons. Jibu & Yasue (1995) proposed another theory according to which quantum fields operate across the entire brain. However, perhaps the most prominent quantum brain theory is the Penrose - Hameroff Orchestrated Objective Reduction (Orch OR) theory (Hammeroff, 1998; Hameroff & Penrose, 2014).

Using Gödels incompleteness theorem as a case for argument, Penrose (1989) raised the issue that humans are able to generate ideas beyond the capability of classical computer algorithms. He then proposed that quantum computational processes (superposition unitary evolution, and quantum collapse) may be required to realize this capability. The OR part of Orch OR theory refers to the initial proposal by Penrose (1989) that the collapse of superposed brain states is objectively caused by gravitational differences in mass displacements of a superposed state, and this objective collapse produces conscious experience. Thus contrary to previous proposals that consciousness collapses the superposition state, the ORCH OR theory proposes that objective collapse of a superposed brain state causes conscious experience. These initial ideas of Penrose were criticized by physicists on the basis of decoherence – the brain is too wet, warm, and noisy to maintain coherence long enough to compute useful cognitive computations (Tegmark, 2000).

At this impasse, Hameroff, an anesthesiologist, stepped in and identified a possible solution to Penrose's decoherence problem by pointing to the idea that these quantum computations may take place inside the microtubules that form the skeleton within neural cell bodies. Microtubulin quantum states can be initially organized or "orchestrated" by synaptic inputs (the Orch part of the theory). Hameroff argued that microtubule quantum pathways can be protected within the microtubule from decoherence for meaningful times up to 25 msec (see Hagan et al., 2002). Hameroff also proposed the idea that gap junctions between neurons could provide lateral connections among neurons to form a brain wide quantum state. Schrödinger evolution of this brain state could perform quantum computing operations, generating entanglement that could also "orchestrate" the quantum computations. The collapse of the entangled superposition state could then regulate neural synaptic activity and produce conscious experience.

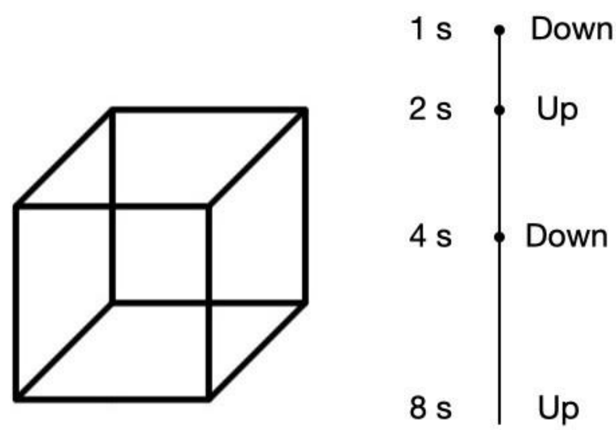
More recently Fisher (2015) (see also Weingarten et al., 2016; Halpern & Crosson, 2019) proposed an alternative solution to the quantum brain decoherence problem. Fisher pointed out that Posner molecules in presynaptic neurons contain phosphate ions that can serve as entangled qubits. Posner molecules can protect their coherent states against fast decoherence (resulting in extreme decoherence times in the range of hours or even days). The entangled phosphates can be transported using vesicular glutamate transporter to different neurons, which can stimulate glutamate release and amplify postsynaptic firing among multiple neurons.

While the issue of issue of decoherence is still being debated, another criticism of the foregoing quantum brain theories is the theoretical gap between the purported quantum brain processes and actual cognitive processing and conscious experience. How can these quantum brain mechanisms scale up to account for basic cognitive processes like memory, categorization, judgment, and decision making? How can they account for the contents of our conscious experiences? Perhaps quantum cognition can help reduce this theoretical gap.

Quantum Cognition

Quantum cognition concerns the application of the mathematical principles of quantum theory to human behavior, but without the physics. Using the same mathematics entails a number of consequences; one of which is that quantum cognition shares an important problem with quantum physics – a measurement problem. However, quantum cognition faces a different kind of measurement problem called *self-measurement*. Measurement changes the system under study, whether it be an electron or a human, so it is crucial to know when a measurement occurs. It is usually clear when a measurement occurs in physics, for example, measuring the spin of an electron by applying a magnet. In quantum cognition, it is usually assumed that a measurement occurs when the experimenter asks the participate to

Figure 3. Necker cube ambiguous figure on left, record of person's orientation perception on right



respond to a question. But the participant doesn't have to wait for an experimenter to ask a question. Instead, a person can spontaneously ask herself a question about an issue, and make up her mind about it, whenever she wants. Self-measurements are unobservable from the point of view of the experimenter.

Consider, for example, an experiment designed to investigate a person's perception of the orientation of a Necker cube (see left panel of Figure 3). Most people experience the face of the cube flipping from facing up to facing down across viewing time. The experimenter can ask the person to report the orientation whenever the cube changes orientation. But this requires the person to take a series of self- measurements on the perceived orientation of the cube. On the right side of the cube is an example of hypothetical data that is recorded in such an experiment, showing the time points when the orientation was perceived to change. Implicit self-measurements by the person viewing the cube can occur in between the reported times, sometimes producing no change (thus no report) and sometimes producing a change (which is then reported). The problem for building a quantum cognition model of this data is that the theorist does not know when the person is taking self measurements.

Before addressing this self-measurement problem, let us first describe why quantum theory might be useful for modeling human judgment and decision making behavior. The first principle is superposition. Like an electron before measurement, humans can enter states of uncertainty, during which time no definite answer exists in the person's mind, and all answers have some potential to be reported. For example, the person viewing the Necker cube can enter a state of uncertainty in which the person is superposed between up and down orientations at some moment. The second principle is sensitivity to measurement. Like an electron, measurement can change the state of a person from a superposed state to one consistent with the observed answer. For example, upon a self-measurement resulting in the decision to report that the cube is perceived in the up direction, the person's uncertainty is resolved, the superposition collapses, and the person becomes clear that the cube is facing up at that moment. The third principle is non-commutativity. Like an electron, the order of measurements can, but not always, matter for human judgements and decisions. For example, asking the person how many polygons are shown in the image (7 are drawn in the 2-dimensional plane), and then asking whether the image shows a cube facing up or down, could give different answers depending on the order of these questions. The fourth principle is unitary evolution. Like an electron, the state of the person changes and evolves across time until the next measurement occurs. For example, after deciding that the cube is orientated in the up direction, unitary evolution can move the state from the up state to a superposition between up and down, returning the person to a state of uncertainty.

What is the empirical evidence for applying these principles to human judgment and decision making? Like the double-slit experiments in physics that are used to reveal interference effects, experiments on human decision making have also revealed interference effects. For example, consider a study on the prisoner dilemma game by Shafir & Tversky (1992). Briefly, two players have two choices (defect, cooperate), and the payoffs for each person depend on the pair of choices made by each player. However, the payoffs are arranged so that no matter what the opponent does, the player is better off by defecting, so that defection is the dominant choice. However, the payoffs to each player are better if they both cooperate as compared to both defecting. Usually the game is played with both players moving simultaneously, without knowledge of their opponents move. However, Shafir and Tversky introduced a manipulation like the two-slit experiment: in one condition, a participant was informed that the opponent already defected (analogous to closing off one slit in the two-slit experiment), and under another condition a participant was informed that the opponent cooperated (analogous to closing off the other slit in the two-slit experiment). They found 97% defection when the opponent was known to defect, 84% defection when the person was known to cooperate, but only 64% defection when the opponent's move was unknown. An interference effect occurred because the percentage of defection in the unknown case falls far below the total probability formed by any weighted average of the two known cases. Many participants decided to defect when the opponent's action was known, but then switched and decided to cooperate when the opponent's move was unknown. There are many more experiments that have found these types of interference effects (see Pothos & Busemeyer, 2022 for a review).

Quantum cognition models account for these results as follows (see, e.g., Pothos & Busemeyer, 2009). The player's state is represented by a vector $|\psi\rangle$ in an N−dimensional vector space. A projector P_{OD} projects the state on to the subspace corresponding to the event that the opponent defects, and a projector $(I - P_{OD})$ projects on the complement event that the opponent cooperates (the identity I projects on the entire vector space). The event that the player chooses to defect is represented by another projector P_{PD}. Under the usual condition when the player does not know the opponent's move, the probability that the player defects is determined by projecting the state on to the event that the player chooses to defect, and taking the squared magnitude

$$p(PD) = \|P_{PD} \cdot |\psi\rangle\|^2 . \quad (1)$$

The projection that the player defects, located inside the brackets, can be decomposed into a superposition between the opponent defecting and collaborating follows

$$P_{PD} \cdot |\psi\rangle = P_{PD} \cdot (P_{OD} + (I - P_{OD})) \cdot |\psi\rangle \tag{2}$$
$$= P_{PD} \cdot P_{OD} \cdot |\psi\rangle + P_{PD} \cdot (I - P_{OD}) \cdot |\psi\rangle.$$

The squared length of the sum produces the result

$$p(PD) = \|P_{PD} \cdot P_{OD} \cdot |\psi\rangle\|^2 + \|P_{PD} \cdot (I - P_{OD}) \cdot |\psi\rangle\|^2 + Int \tag{3}$$

where the interference term,

$$Int = \mathcal{R}[\langle\psi| P_{OD} \cdot P_{PD} \cdot (I - P_{OD}) \cdot |\psi\rangle],$$

is the crossproduct of the terms being summed in Eq 2. The first term $\|P_{PD} \cdot P_{OD} \cdot |\psi\rangle\|^2$ corresponds to the probability of predicting the opponent defects and then player defects; the second term $\|P_{PD} \cdot (I - P_{OD}) \cdot |\psi\rangle\|^2$ corresponds to the probability of predicting the opponent cooperates and then player defects; and their sum corresponds to the classical total probability. The third term is the interference, which produces violations of the classical law of total probability. If the projectors, P_{PD} and P_{OD} commute then the interference is zero. If they don't commute, then this interference term can be negative and lower the probability to defect in the unknown case below both of the probabilities when the opponent's move is known, thus accounting for the experimental results. For general introductions to quantum cognition, see A. Y. Khrennikov (2010) and Busemeyer & Bruza (2012).

Connections

What do these three topics have to do with each other? Quantum theories of consciousness have more to say about quantum physics than cognitive psychology and conscious experiences. Quantum brain theories have not been sufficiently scaled-up to provide clear implications for how quantum physical processes generate more complex cognition. Quantum cognition theories have avoided addressing fundamental issues about consciousness and have remained agnostic with respect to the quantum brain hypothesis. Now we address the problem of connecting these ideas together by using quantum cognition to make bridges across the other two topics.

Quantum brains and quantum cognition

If we put aside the debate about decoherence for sake of discussion, and consider the Orch OR quantum brain theory, then how can quantum cognition form a *bridge* between Orch OR theory and judgment and decision making behavior? One bridge that can be made is to the phenomena of question order effects (Hameroff, 2013). For example, returning to the prisoner dilemma task, participants are more likely to defect if they are asked to predict their opponent's move first and then decide their own action as compared to the reverse order (Tesar, 2020). Quantum cognition models account for question order effects by using what are called non-commuting projectors to represent the answers to different questions. Non commuting projectors are constructed from unitary transformations of the basis used to describe each question.[6]

Hameroff (2013) suggested a way to implement the question order model from quantum cognition in the Orch OR theory as follows. For simplicity let us consider the simple two dimensional quantum model for the prisoner dilemma game used by Tesar (2020), which is model is analogous to that used for measuring the spin of a single electron in two different orientations. In this case there are two binary valued (defect/cooperate) questions: one asking the person to predict the opponent's move, and the other asking for the person's own decision. Hameroff suggested that tubulins inside a microtubule form quantum channels, which can be orientated in one of two directions. In this example, one orientation could represent the prediction that the opponent will defect, and the other could represent the prediction that the opponent will cooperate. Furthermore, the quantum channel pathways can become superposed to form a quantum qubit $|\psi\rangle = \psi_D \cdot |pred\ Defect\rangle + \psi_C \cdot |pred\ Coop\rangle$,

[6]Unitary transformations are generated in real time by the Schrödinger differential equation.

where ψ_D is the potential that defect is predicted, and ψ_C is the potential that cooperate is predicted. The OR process would then collapse on either the prediction for defection with probability $|\psi_D|^2$ or cooperation with probability $|\psi_C|^2$.

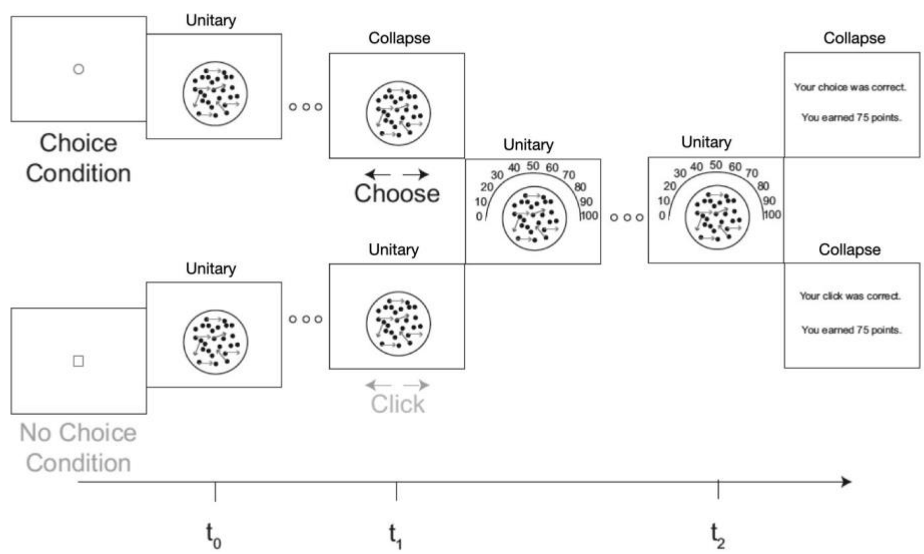
So far, this matches a quantum cognition $2-d$ model for the first question about what prediction to make. However, what is missing in this account is the brain mechanism that changes the basis needed to answer the second question about the person's own decision.[7] To produce question order effects, a unitary transformation U is needed to change the potentials for the prediction basis (ψ_D, ψ_C) to potentials for decision basis (φ_D, φ_C) to produce a new superposition: $|\psi\rangle = \varphi_D \cdot |decide\ Defect\rangle + \varphi_C \cdot |decide\ Coop\rangle$. Note that the same qubit, $|\psi\rangle$, is used for both questions, and only the basis used to evaluate this qubit is changed. This is one way that quantum cognition theory provides a bridge by scaling up quantum brain mechanisms to account for question order effects.

Another bridge that can be made is for dynamic interference effects (Hameroff, 2014). For example, Kvam et al. (2015) examined the interference effects of choice on later confidence ratings (see Figure 4). This experiment is also analogous to a double slit experiment in physics. Participants were asked to view a circle of randomly jiggling dots, and while most of the dots jiggled randomly, a small percentage (e.g., 5%) jiggled in a systematic direction (e.g., right). The participants were asked to either make a choice (decide whether the dots were primarily moving to the left or the right) or rate the probability that they were moving to the right. The experiment involved 2 conditions: a choice-rating condition (shown as the top sequence in Figure 4) and a rating-only condition (shown as the bottom sequence in the figure). Referring to Figure 4, the dot display appeared at time t_0, choice was required at time t_1 for choice-rating condition (but not for the rating-only condition), and then the dot display remained on until time t_2 at which point in time the probability ratings were made for both conditions.

The experiment was designed to test quantum versus Markov random walk models of evidence accumulation. The critical test concerns the marginal distribution of probability

[7] Hameroff (2013) suggested that the two questions could be represented by a two qubit entangled state, but that representation produces commutative measurements and would not produce question order effects.

Figure 4. Random dot motion experiment. Unitary evolution occurs until a measurement is made.



ratings at time t_2 (pooled across choices for the choice - rating condition). The Markov walk model assumes that a person's beliefs follow a classical trajectory like a particle, so that the belief is precisely located at each point in time, and the choice simply records whether the existing location lies above or below a threshold. Consequently, it predicts no effect of choice on the marginal probabilities, and thus no difference between conditions.[8] In contrast, the quantum walk model assumes that beliefs are superposed across the probability scale at each point in time, producing a wave that flows across time. But choice collapses the wave to one side of the probability scale, which changes the interference pattern, producing a difference between the conditions. In support of the quantum walk model and contrary to the Markov random walk model, the experiment produced systematic differences between the conditions.

Hameroff (2014) proposed a way to represent these kinds of quantum walks in the Orch OR theory. He suggested that the lattice forming the basis states of the quantum

[8]The formal mathematical derivation appears in Kvam et al. (2015).

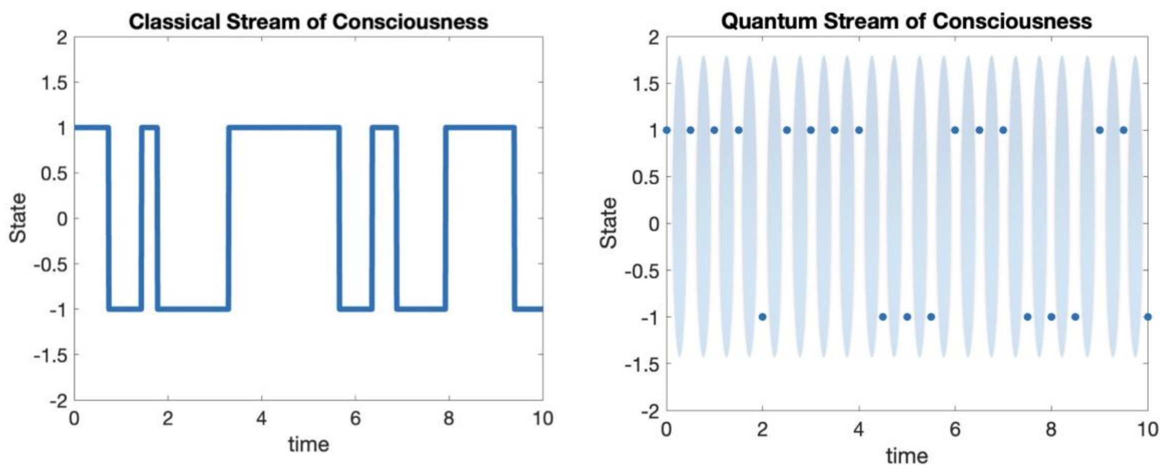
walk can be mapped into "topological qubits" in brain neuronal microtubules. According to Hameroff, the OR process picks one of the possible paths through the quantum walk lattice to produce the choice probability by squaring the magnitude of the path amplitude. The quantum walks used in quantum cognition (Kvam et al., 2015; Fuss & Navarro, 2013) produce the choice probabilities differently. Before measurement, the state of the quantum walk evolves unitarily across time, which in Feynman's view, can be interpreted as traveling all paths in parallel. The probability of a choice is computed by projecting the current state of the quantum walk on a subspace corresponding to the choice. In Feynman's view, this measurement first sums across paths that lead to the subspace, and then squares the length of the sum of paths. This is another way that quantum cognition theory provides a bridge by scaling up quantum brain mechanisms to account for dynamic interference effects.

Whether quantum processes occur within microtubules or Posner molecules, the computations of these quantum brain theories are assumed to be based on quantum computing operations applied to entangled qubits. So far, quantum cognition models have not been limited to using computing operations based on entangled qubits. However, it is not difficult to re-code an $N-$ dimensional quantum cognition model into a 2^n entangled qubit model with a sufficiently large number n of qubits. Recently, in fact, the quantum walk models used in quantum cognition have been implemented using entangled qubits on quantum computers (Pothukuchi et al., 2023).

Quantum consciousness and quantum cognition

Almost all of the past applications of quantum cognition theory have been focused on predicting human behavior (however, see some recent attempts to address consciousness described later in this article). So the issue of when consciousness occurs has generally been avoided. However, if self-measurement is connected to consciousness experience, then this becomes an important theoretical issue for quantum cognition. Consider the theoretical assumptions that Kvam et al. (2015) used to model the random dot motion experiment shown Figure 4. Their computations were based on assuming that unitary evolution of a

Figure 5. classic trajectory and quantum stream



superposition state always occurred except when the experimenter explicitly requested a measurement (a report of a choice or a rating), at which point a projection (collapse) would occur. This assumption is shown in Figure 4 by the labels "unitary" and "collapse" shown at various time points in the figure. If instead, self-measurements and conscious experiences occurred implicitly at several places in between the experimentally requested measurements, then this would drastically change the predicted interference patterns. For example, if the participants in the rating-only condition made an implicit choice around the same time as t_1, then this would have eliminated any difference between the conditions. The fact that a difference did occur suggests that their assumptions were reasonable, but this remains an open question requiring further research.

What can quantum cognition contribute to understanding consciousness? One answer to this question concerns the implications for stream of consciousness across time. Consider once again a participant who is asked to report the orientation of the cube whenever it changes across time. Figure 5 shows two different types of dynamic laws for consciousness: classical dynamics on the left, and quantum dynamics on the right. According to classical dynamics, the person's conscious perception is always precisely located. In the left panel, the perception is located either at the right (coded +1) or on the left (coded -1).[9] For example, it

[9]Two definite states are shown for simplicity, but there could be more. The point is that the person is

is located at -1 at time 2 but switches to +1 before time 4. Consequently, if asked about the location at some point in time (e.g., time 4), the person simply reports the existing location (+1). According to quantum dynamics, the person's conscious perception is located precisely only at time points where a self-measurement occurs (indicated by dots), and otherwise it is dispersed (indicated by a smeared out image) across states in a superposition. Before the self-measurement, a precise location doesn't exist, and afterwards, the self-measurement creates a precise location out of the superposition state.

How can one experimentally test which of the two dynamic laws applies to the stream of consciousness? The dynamic interference effects, discussed earlier by Kvam et al. (2015), provide some initial evidence for the quantum dynamic view. Below we describe two other more general ways to investigate this question.

Perhaps the strongest possible test is based on the following temporal Bell inequality experimental design (Leggett & Garg, 1985; Suppes & Zanotti, 1981). This experiment is a dynamic version of the Bell inequality experiment, which is one of the most famous experiments in physics.[10] The original Bell experiment involved 4 different conditions, but the temporal Bell only involves three conditions. Consider an experimental test applied to the Necker cube ambiguous perception phenomena (Atmanspacher & Filk, 2010)

For this design, the experiment involves 3 time points; however, the participant is not asked the orientations at each time point. Instead, only a pair of time points is compared. Under condition C_1 the participant is asked whether or not the cube changed from time t_1 to t_2; under condition C_2 the participant is asked whether or not the cube changed from time t_2 to t_3; and under condition C_3 the participant is asked whether or not the cube changed from time t_1 to t_3. According to a classical "trajectory" theory, each trial should pass through one of the rows shown in Table 1 under the first three columns labelled S_1, S_2, S_3 at

precisely located in one of the definite states.

[10]The Bell experiment has a long history. An experiment was originally proposed in the 1930's by Einstein, Pudowsky, and Rosen, but it was difficult to realize because it involved position and momentum measurements. In the 1950's, Bohm redesigned the experiment using two valued spin of electrons, which made the experiment more feasible. The inequality was originally proposed in the 1960's by John Bell but not experimentally tested. Actual experimental tests were later conducted by John Clauser first in the 1970's, then Alain Aspect in the 1980's, and most recently Anton Zielinger. The latter three were recently awarded Nobel prizes in physics for this work.

Table 1: Temporal Bell experimental paradigm and predictions of a classical trajectory

Trajectory	S_1	S_2	S_3	C_1	C_2	C_3
1	+	+	+	0	0	0
2	+	+	-	0	1	1
3	+	-	+	1	1	0
4	+	-	-	1	0	1
5	-	+	+	1	0	1
6	-	+	-	1	1	0
7	-	-	+	0	1	1
8	-	-	-	0	0	0

The values assigned to the row under S_1, S_2, S_3 indicate the states at the three experimental time point that occur with a classical trajectory. Condition C_1 tests change from t_1 to t_2; C_2 tests change from t_2 to t_3; C_3 tests change from t_1 to t_3. 0 = no change, 1 = change.

the three experimental tested time points. For example, the trajectory shown on the left in Figure 5 at times 2,4, and 6 would pass through row 6, which produces $C_1 = 1$ to indicate a change from 2 to 4 seconds, $C_2 = 1$ to indicate a change from 4 to 6 seconds, and $C_3 = 0$ to indicate no change occurred for the 2 to 6 second pair. On each trial, the participant's classical perception must pass through one of the 8 possible trajectories. Note that whenever conditions $C_3 = 1$, either condition $C_1 = 1$ or $C_2 = 1$, but there are cases when $C_3 = 0$ yet either $C_1 = 1$ or $C_2 = 1$ can occur. Therefore, the classical trajectories must obey the probabilistic inequality $p(C_1 = 1) + p(C_2 = 1) \geq p(C_3 = 1)$. Only a few experiments have been conducted to test this inequality, but some preliminary studies suggest that it can be violated (Waddup et al., 2023; Asano et al., 2014).

Quantum dynamics can violate the temporal Bell inequality. For example, Atmanspacher & Filk (2010) proposed a 2-dimensional quantum oscillator model for the Necker cube task, and their predictions, calculated using empirically estimated parameters, result in violations of the inequality. However, attempts to model the temporal Bell experiments also encounter self- measurement problems, because it is assumed that measurements only occur at the experimentally programmed time points. As noted before, self-measurements could confound the programmed experimental measurements.

Another interesting application to stream of consciousness concerns the quantum zeno effect, named after the fifth century B.C. Greek philosopher, Zeno of Elea, who proposed the "arrow" paradox (if an arrow in flight was watched at every instant, it would never seem to move). According to quantum theory, rapidly repeated measurements, producing repeated collapse, can slow down and freeze the state from changing (Misra & Sudarshan, 1977).

Atmanspacher et al. (2004) examined the effect of repeated measurements predicted by their 2-dimensional oscillator model for the Necker cube. Using empirically estimated parameters, their calculations from the 2-dimensional oscillator model produces a "quantum zeno" slow down in the rate of change in perceived orientation of the cube. Yearsley & Pothos (2016) experimentally tested for a zeno effect using a decision making task involving the presentation of evidence about a crime. Using the same sequence of evidence, they manipulated the frequency that a participant made explicit judgments about the crime, and they found a systematic slow down of opinion change with increased frequency of measurements. Once again, experimental control over the frequency of measurement may be confounded by the possibility of self-measurements. The fact that the frequency manipulation succeeded to produce a zeno effect suggests that self-measurements did not occur often.

The psychological importance of self-measurements was first discussed by Stapp (1993/2009). He built on the earlier idea about volition by William James, who proposed that a person can control one's own actions by consciously attending to a plan for the action. Stapp interpreted conscious attention as repeated self-measurements. He proposed that repeated self-measurements can produce repeated quantum collapse, and consequently induce a quantum zeno effect effect in the brain. Assuming that a person has free choice of what and when to measure, a person can apply willful effort to repeatedly apply measurements that hold an intention for action in place in the face of disturbances.

Future questions

A large literature has already grown to a fairly mature and sophisticated level in the fields of quantum consciousness, quantum brains, and quantum cognition, but what has been missing in all of this work is a systematic way to connect these three fields together.

To close this gap, we have used quantum cognition to make bridges between the other two. Although we have discussed some important connections in the previous sections, this is just a beginning and there are many more important questions remain to be addressed. Below we outline two important directions for future work.

When does consciousness occur and what become conscious?

Although past quantum cognition research has focused mainly on judgment and decision behavior, some recent efforts have been made toward applications to consciousness (Tsuchiya et al., 2024; A. Khrennikov, 2015, 2023). These previous works generally agree on the principle that consciousness is generated by some type of measurement. The measurement could be self-generated (questioning oneself) or externally generated by the environment (another person asking a question). Note that there can't be any collapse to form an experience without first selecting a measurement, because the measurement determines the basis for evaluation, and without a basis, there is nothing singled out to collapse to.

At this point it is useful to briefly review developments in quantum measurement theory. A measurement has a basis that is formed by a collection of basis vectors (eigenvectors) that are used to describe all of the outcomes of the measurement. Von Neumann's initial measurement theory assumed a collapse to a single dimensional ray spanned by one of these basis vectors. Later, Lüder generalized the measurement theory by allowing collapse to a multi-dimensional subspace rather than a single dimensional ray. This early theory has been substantially modified in later times. In modern quantum measurement theory, it is useful to formulate the entire quantum state as combination of a system state of interest (e.g., the electron or a person's opinion) and an environment state (the measurement instrument or the experimenters question). The system may start out independent of the environment before their interaction, but after their interaction, the measurement becomes correlated with the system. The interaction is generated by a unitary operator that changes the entire quantum state from one superposition state to another over time (see Appendix for an example). Thus the state after the interaction is still a superposition, but it can now be decomposed in terms of the basis vectors that describe the measurement.

According to collapse theories, the decomposed superposition state, collapses, in a non - unitary and probabilistic manner, to one of the possible measurement outcomes. When considering conscious states, the vector space must be very high dimensional feature space. A basis vector used to describe the space can represent a large combination of feature values. Thus, the superposition state must be very high dimensional, and a measurement usually is a question only about a very few features included in this very high dimensional state. Given the high dimension of the vector space, the decomposition must partition this huge vector space into massive subspaces. The collapse is a projection of the superposition state on to one these massive subspaces. Therefore, the collapse must retain a very large amount of the the original superposition state. Consider once again the Schrödinger cat paradox. The superposition state was decomposed into a superposition over the outcomes of the cat being dead or cat alive. However, the cat has many features, like its size, color, type, position, face, eyes, ect. The collapse on the cat is dead outcome, for example, only changes the feature about it being alive or not, and the projection would still contain all the information about the myriad other features of the cat before the collapse. Thus only a very small part of the decomposed state is lost in the projection.

The previous works on quantum cognition generally agree that the evolution of a superposition state, before a measurement is selected for evaluation, is not consciously experienced (unconscious). As Stapp has emphasized, this is because, although the state is ready to be evaluated by some measurement, before the measurement is selected, there is no basis selected for decomposing the superposition. The same superposition can be decomposed in many ways. There is nothing specific or unique upon which a person can be aware until a measurement is chosen that provides the basis for measurement.

Let us take a concrete example. Suppose that while driving a car down a road, a cat runs across the road. The driver's physical visual system processes a scene of some kind of small animal running across the road. The current working memory mental state (which may also be neural or not) is updated with this new information, in which case the driver remains uncertain about the situation. At this point, she could ask herself - what kind of animal was that? Or she could ask herself - should I slam on the brakes? Or she could ask herself- am I late for my meeting? There are very many questions she could be considering at that time. Or maybe a friend in the car could ask a question. Using this example, we face the issue - when does consciousness begin and what is its contents?

According to a quantum cognition theory, the visual processing of the scene organizes some mental superposition state, and the question (measurement) applied to this state forms the basis for evaluating the state. In terms of a system and environment representation, the visual process generates the system part, and the person's questions about the scene represent the measurement part. Once the basis is selected, the state can be decomposed with respect to its basis vectors. If the driver asks the "what is it" question, the basis will be formed from different kinds of small animals; If she asks the "what should I do" question, the basis will be formed from different kinds of actions, and so on. Finally some projection on a subspace spanned by the basis vectors is chosen (probabilistically according to quantum rules) to generate a reported answer. For example, if the passenger asks "what is it ", then the driver might verbally answer "cat" . The standard quantum consciousness theory answer to the question "when does consciousness begin" is that it begins upon projection, that is the collapse. In other words, this is when the person consciously perceives the answer - the person thinks she saw a cat.

Other possible solution to the question "when does consciousness occur" have been proposed. Neven et al. (2024), following Everett's theory, propose that collapse/projection never occurs, and there is only superposition. Consciousness occurs when the superposition state is decomposed by selection of a measurement. The problem is that this does not yet explain why we seem to experience only one outcome of the decomposition. To do this, Neven et al. 2024 adds the assumption that "many worlds" or "many minds" are created by the decomposed superposition produced by the measurement, and one our conscious experiences (probabilistically according to quantum rules) lands in one of them. For example, if the "what is it" question is asked, then one "mind" experiences the cat, another experiences a dog, another experiences a squirrel, and so on. The mind of the driver that has landed in the "cat" world verbally reports cat to the passenger.

There is, of course one logically remaining, although perhaps even more controversial hypothesis - suppose we *can* experience a decomposed superposition state after a measurement is selected for evaluation, without collapse, but only to some degree or level of awareness. Evidence comes from the fact that participants in our experiments can report subjective likelihoods on probability rating scales and subjective feelings on attitude scales. Considering our example, the driver may be able to vaguely feel some potential that it was a cat, or dog, or squirrel, or something else. The potentials assigned to a decomposed superposition state could grow by unitary (or more elaborate Lindblad to stabilize the state) evolution over time, so that one set of potentials are collectively driven toward unity, and the remainder toward zero, producing a growing awareness of one outcome (e.g., one animal such as the cat). The potentials of a decomposition can determine the prominence or salience of that outcome in our conscious experience.

This hypothesis seems to go against one of the main axioms of many consciousness theories – the unity of consciousness. According to this controversial "superposed consciousness" hypothesis there is no quantum measurement problem, instead there is a problem with one of the axioms of consciousness theories. However, in agreement with the unity axiom, it may impossible to overtly express a superposition state to the outside world, and so we cannot directly communicate to others our superposition state (we can't copy our quantum state). Even a probability rating, based on a decomposed superposition state, is actually an outcome of a measurement. In order to communicate external information about a purported consciously experienced superposition, we may need to choose a subspace to report probabilistically according to quantum rules. But this overt report may only produce some back action on the superposition state generated by a unitary operator, and not necessarily produce a complete collapse (loss of information) of the superposition state.

How to select the measurement or basis for evaluation?

An important issue raised by Stapp (1993/2009) concerned how a person selects what to measure and when to measure. In other words, what determines the basis for evaluating and decomposing the superposition state from moment to moment? Stapp identified this process for selecting a measurement as a new type 0 process to be added before Von Neumann's type I and II processes. He argued that the selection is under the free will of a person. Although he argued that this selection process stands outside physical theory, he did link the selections to attention theories, which are very thoroughly developed in cognition (Pashler, 1998; Wickens et al., 2022; Tsotsos, 2021).

Collapse theories seem to have difficulty accounting for the stream of consciousness. If consciousness first requires a measurement to be selected in order to decompose the superposition state, and a collapse on the basis formed by this measurement is required, then conscious experience would be very discreet, as shown in the right hand panel of Figure 5. Of course, like rapid frames in a movie, this process could occur very rapidly and appear continuous. Alternatively, if no collapse ever occurred, and the superposition state could be consciously experienced, then it is not necessary to assume any punctuated experiences of consciousness (remove the dots in Figure 5). Instead, there would be smooth and continuous evolution of conscious experience of a decomposed superposition. The decomposition itself can be produced by a continuously varying unitary operator that forms the interaction between the system and the environment.

Chalmers & McQueen (2022) addressed the basis selection problem differently. Rather than have the basis change across time under the free will control of a person, they assumed that the basis for consciousness was fixed. They defined this fixed basis as the set of all the possible total states of consciousness, which, for concreteness, they assumed were defined by the states of information integration theory (e.g., Tononi, 2017). They also proposed that a superposition of physical states in the environment gets entangled and correlated with a superposition of neural states within a person, and the latter is psychophysically mapped to a corresponding superposition of total conscious states. However, the total conscious states are hypothesized to be super resistant to superposition, so they rapidly and spontaneously collapse to a single total conscious state, which then by entanglement with neural and physical state, cause them to collapse too.

From the point of view of quantum cognition, a fixed basis for the decomposition of superposition states faces problems. The quantum probabilities produced by a fixed basis always agree with classical probabilities. The basis needs to change in order to account for empirical findings from judgment and decision making, such as interference effects, question order effects, and other context effects. These effects are explained in quantum cognition by evaluating the same superposition state using a different measurement basis for decomposition with each question.

Concluding comments

The development of science cannot be determined only by people's personal will. When we are on the road of scientific research, we need to stop and think for a while in order to have a clearer direction of progress. 100 years ago, when people were faced with some experimental results that were contrary to the laws of classical physics, scientists such as Bohr, Heisenberg, Schrödinger, and others pieced together quantum theory. Now, the same situation is happening again in the realm of consciousness, and quantum theory may once again exert its huge power. It provides a new perspective for people to comprehensively understand consciousness, judgment, and decision-making, and it may also herald the arrival of a scientific revolution.

References

Asano, M., Hashimoto, T., Khrennikov, A., Ohya, M., & Tanaka, Y. (2014). Violation of contextual generalization of the leggett–garg inequality for recognition of ambiguous figures. *Physica Scripta*, *2014* (T163), 014006.

Atmanspacher, H. (2011). Quantum approaches to consciousness. In (p. x). Stanford Encyclopedia of Philosophy.

Atmanspacher, H., & Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. *Journal of Mathematical Psychology*, *54* , 314-321.

Atmanspacher, H., Filk, T., & Romer, H. (2004). Quantum zero features of bistable perception. *Biological Cybernetics*, *90* , 33-40.

Beck, F., & Eccles, J. (1992). Quantum apsects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences of the USA*, *89* , 11357-11361.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambirdge University Press.

Chalmers, D. J., & McQueen, K. J. (2022). Consciousness and quantum mechanics. In S. Gao (Ed.), (p. 11-63). Oxford University Press.

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

d'Espagnat, B. (2005). Consciousness and the wigner's friend problem. *Foundations of Physics*, *35* , 1943–1966.

Fisher, M. P. (2015). Quantum cognition: The possibility of processing with nuclear spins in the brain. *Annals of Physics*, *362* , 593–602.

Fuss, I. G., & Navarro, D. J. (2013). Open parallel cooperative and competitive decsision processes: A potential provenance for quantum probability decision models. *Topics in Co*, *5 (4)*, 818-843.

Gao, S. (2008). A quantum theory of consciousness. *Mind and matter*, *18* , 39-52.

Hagan, S., Hammeroff, S., & Tuszynksi, J. (2002). Quantum computation in brain microtubles? decoherence and biological feasibility. *Physical Reviews E* , *65* , 061901.

Halpern, N. Y., & Crosson, E. (2019). Quantum information in the posner model of quantum cognition. *Annals of Physics*, *407* , 92–147.

Hameroff, S. (2013). Quantum mechanical cognition requires quantum brain biology. *Behaviora and brain sciences*, *36 (3)*, 287-288.

Hameroff, S. (2014). Quantum walks in brain molecules – a biomecular basis for quantum cogntion. *Topics in cognitive science*, *6*, 91-97.

Hameroff, S., & Penrose, R. (2014). Conscoiusness in the universe: A review of the orch or theory. *Physics of life reviews*, *11*, 39-78.

Hammeroff, S. (1998). Quantum computation in brain microtubles? the penrose - hameroff "orch or" model of consiousness. *Philosophical Transactions Royal Society London (A)*, *356*, 1869-1896.

Jibu, M., & Yasue, K. (1995). *Quantum brain dynamics and consiousness*. Amsterdam: Benjamins.

Khrennikov, A. (2015). Quantum like model of unconscious - conscious dynamics. *Frontiers in psychology*, *6*, 997-1010.

Khrennikov, A. (2023). *Open quantum systems in biology, cognitive , and social sciences*. Springer.

Khrennikov, A. Y. (2010). *Ubiquitous quantum structure: From psychology to finance*. Springer.

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence. , 112 (34). *Proceedings of the National Academy of Science*, *112 (34)*, 10645-10650.

Leggett, A. J., & Garg, A. (1985). Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks. *Physical Review Letters*, *54*, 857-860.

Misra, B., & Sudarshan, E. (1977). The zenos paradox in quantum theory. *Journal of Mathematical Physics*, *18*, 756-763.

Neven, H., Zalcman, A., Read, P., Kosik, K. S., van der Molen, T., Bouwmeester, D., . . . Koch, C. (2024). Testing the conjecture that quantum processes create conscious experience. *Entropy*, *26* (6), 460.

Nielsen, M. A., & Chuang, I. L. (2000). *Quantum computation and quantum information.* Cambridge University Press.

Pashler, H. (1998). *The psychology of attention*. MIT Press.

Penrose, W. (1989). *The emperor's new mind*. Oxford University Press.

Plotnitsky, A. (2012). *Niels bohr and complementarity: An introduction*. Springer Science & Business Media.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of 'rational' decision making. *Proceedings of the Royal Society, B.*, *276 (1665)*, 2171-2178.

Pothos, E. M., & Busemeyer, J. R. (2022). Quantum cognition. *Annual review of psychology*, *73* , 749–778.

Pothukuchi, R. P., Lufkin, L., Shen, Y. J., Simon, A., Trevisan, B. E., Tu, M., . . . others (2023). Quantum cognitive modeling: New applications and systems research directions. *arXiv preprint arXiv:2309.00597* .

Pylkkannan, P. (2006). *Mind , matter and the implicate order*. Springer.

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews: Neuroscience*, *23*, 439-452.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: nonconsequential reasoning and choice. *Cognitive Psychology*, *24* , 449-474.

Stapp, H. (1993/2009). *Mind, matter, and quantum mechanics* (3rd ed.). Springer.

Suppes, P., & Zanotti, M. (1981). When are probabilsitic explanations probable? *Synthese*, *48* (191-199).

Tegmark, M. (2000). Importance of quantum decoherence in brain processes. *Physical Review E*, *61 (4)*, 4194-4206.

Tesar, J. (2020). A quantum model of strategic decision-making explains the disjunction effect in the prisoner's dilemma game. *Decision*, *7* ((1)), 43-54.

Tononi, G. (2017). The blackwell companion to consciousness. In S. Schneider & M. Velmans (Eds.), (p. 243-257). Wiley Blackwell.

Tsotsos, J. K. (2021). *A computational perspective on visual attention*. MIT Press.

Tsuchiya, N., Bruza, P., Yamada, M., Saigo, H., & Pothos, E. (2024). Quantuml-like qualia hypothesis: from quantum cognition to quantum perception. *Frontiers in Psychology*.

Von Neumann, J. (1932/1955). *Mathematical foundations of quantum theory*. Princeton University Press.

Waddup, O. J., Yearsley, J. M., Blasiak, P., & Pothos, E. M. (2023). Temporal bell inequalities in cognition. *Psychonomic Bulletin & Review*, 1–8.

Weingarten, C. P., Doraiswamy, P. M., & Fisher, M. (2016). A new spin on neural processing: quantum cognition. *Frontiers in human neuroscience*, *10*, 541.

Wickens, C. D., McCarley, J. S., & Gutzwiller, R. S. (2022). *Applied attention theory*. CRC press.

Yearsley, J. M., & Pothos, E. M. (2016). Zeno's paradox in decision-making. *Proceedings of the Royal Society B: Biological Sciences*, *283* (1828), 20160291.

Appendix

This appendix provides an example of using the system plus environment measurement model applied the prisoner dilemma experiment discussed in the article. The experiment is like a double-slit quantum experiment. Under one condition, the player's prediction about the opponent's move is unknown before the player makes a decision (corresponding to leaving both slits open). In the other condition, the player is informed and knows what the opponent will do (corresponding to closing one slit and leaving only one slit open). The interference effect refers to the difference between these two conditions in the probabilities that the player defects. The player's disposition to defect is regarded as the system, and the information about the opponent is the environment.

Define the state of the system plus environment state as

$$|\zeta\rangle = |\psi\rangle \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle)$$

where $|\psi\rangle$ is an $N-$dimensional (possibly very high dimensional) vector in the Hilbert state of the player; and (for simplicity) $|\varphi_O\rangle$ is a vector in a $2-$dimensional environment space spanned by the pair of basis vectors $(|\varphi_{OD}\rangle, |\varphi_{OC}\rangle)$ representing the opponent defects and the opponent cooperates, respectively; and (again for simplicity) $|\varphi_P\rangle$ is a vector in a $2-$dimensional environment space spanned by the pair of basis vectors

($|\varphi_{PD}\rangle$, $|\varphi_{PC}\rangle$), representing player chooses to defect and chooses to cooperates, respectively. Define $I_n = diag\ 1\ 1 \ldots 1$ as a n dimensional identity matrix. The projector P_{OD} projects $|\psi\rangle$ on to the subspace representing the player's belief that the opponent defects, and $P_{OC} = I_N - P_{OD}$; the projector P_{PD} projects $|\psi\rangle$ on to the subspace representing the player's preference to defect, and $P_{PC} = I_N - P_{PD}$.

The state can be decomposed with respect to the prediction measurement as follows

$$|\psi\rangle = P_{OD} \cdot |\psi\rangle + P_{OC} \cdot |\psi\rangle. \tag{4}$$

Alternatively, it can be decomposed with respect to the decision measurement as follows

$$|\psi\rangle = P_{PD} \cdot |\psi\rangle + P_{PC} \cdot |\psi\rangle. \tag{5}$$

It can also be decomposed with respect to both when evaluating prediction first as

$$|\psi\rangle = P_{PD} \cdot P_{OD} |\psi\rangle + P_{PD} \cdot P_{OC} \cdot |\psi\rangle + P_{PC} \cdot P_{OD} |\psi\rangle + P_{PC} \cdot P_{OC} \cdot |\psi\rangle \tag{6}$$

The environmental state and projectors for the measurements can be defined as

$$|\varphi_O\rangle = \alpha_{OD} \cdot |\varphi_{OD}\rangle + \alpha_{OC} \cdot |\varphi_{OC}\rangle$$

$$|\varphi_P\rangle = \alpha_{PD} \cdot |\varphi_{PD}\rangle + \alpha_{PC} \cdot |\varphi_{PC}\rangle\ R_{OD} =$$

$$|\varphi_{OD}\rangle\langle\varphi_{OD}|,\ R_{OC} = I_2 - R_{OD}\ R_{PD} = |\varphi_{PD}\rangle$$

$$\langle\varphi_{PD}|,\ R_{PC} = I_2 - R_{PD}$$

Define the controlled U gate operator that performs the measurement of the player's prediction as

$$U_O \cdot (P_{OD} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle) = (P_{OD} \cdot |\psi\rangle) \otimes (|\varphi_{OP}\rangle \otimes |\varphi_P\rangle)$$

$$U_O \cdot (P_{OC} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle) = (P_{OC} \cdot |\psi\rangle) \otimes (|\varphi_{OC}\rangle \otimes |\varphi_P\rangle)$$

Note that this unitary operator correlates the prediction measurement with the system state. Define the controlled U gate operator that performs the measurement of the player's decision as

$$U_P \cdot (P_{PD} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle) = (P_{PD} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_{PD}\rangle)$$

$$U_P \cdot (P_{PC} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle) = (P_{PC} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_{PC}\rangle).$$

Note that this unitary operator correlates the decision measurement with the system state. Define the projector for player predicts opponent defects as

$$M_{OD} = (I_N \otimes R_{OD} \otimes I_2).$$

Define the projector for player decides to defect as

$$M_{PD} = (I_N \otimes I_2 \otimes R_{PD}).$$

There is no collapse in this model, but we still need to use the projectors to compute the probabilities of the observed responses.

First we consider the evolution of the state when the prediction of the opponent is left unknown, but the player's decision is measured

$$U_P \cdot |\zeta\rangle = U_P \cdot |\psi\rangle \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle)$$

$$= U_P \cdot (P_{PD} \cdot |\psi\rangle + P_{PC} \cdot |\psi\rangle) \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle)$$

$$= U_P \cdot (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_P\rangle) + U_P \cdot (P_{PC} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_P\rangle)$$

$$= (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PD}\rangle) + (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PC}\rangle) .$$

Note that the measurement of the player's decision is a superposition where the state of the player is correlated with the report of the decision.

The probability to defect when no prediction information is provided then equals

$$p(PD) = \| M_{PD} \cdot U_P \cdot |\zeta\rangle \|^2$$

$$= \| M_{PD} \cdot (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PD}\rangle) + M_{PD} \cdot (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PC}\rangle) \|^2$$

$$= \| (I_N \otimes I_2 \otimes R_{PD}) \cdot (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PD}\rangle) + (P_{PD} |\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PC}\rangle) \|^2$$

$$= \| P_{PD} (|\psi\rangle \otimes |\varphi_O\rangle \otimes |\varphi_{PD}\rangle) \|^2$$

$$= \| P_{PD} |\psi\rangle \|^2$$

which is equal to Equation 1, and as can be seen in Equation 3, this probability contains the interference term. In sum, when no prior measurement of prediction occurs, the probabilities do not satisfy the classical law of total probability because interference is present.

Next we consider the evolution of the state when information about the opponent is known and the measurement represented by U_C for prediction about the opponent is applied first and measurement about the player's decision U_P is applied second. To go from the first to the second line, we use the decomposition of both in Equation 6

$$U_P \cdot U_O \cdot |\zeta\rangle = U_P \cdot U_O \cdot |\psi\rangle \otimes (|\varphi_O\rangle \otimes |\varphi_P\rangle)$$

$$= (P_{PD} \cdot P_{OD} \cdot |\psi\rangle \otimes |\varphi_{OD}\rangle \otimes |\varphi_{PD}\rangle)$$

$$+ (P_{PD} \cdot P_{OC} \cdot |\psi\rangle \otimes |\varphi_{OC}\rangle \otimes |\varphi_{PD}\rangle)$$

$$+ (P_{PC} \cdot P_{OD} \cdot |\psi\rangle \otimes |\varphi_{OD}\rangle \otimes |\varphi_{PC}\rangle)$$

$$+ (P_{PC} \cdot P_{OC} \cdot |\psi\rangle \otimes |\varphi_{OC}\rangle \otimes |\varphi_{PC}\rangle).$$

Note that the complete state has been decomposed into a superposition of four mutually exclusive states.

The probability to defect for the known condition then equals

$$p_T(PD) = \|M_{PD} \cdot U_P \cdot U_O \cdot |\zeta\rangle\|^2$$

$$= \|(I_N \otimes R_{OD} \otimes R_{PD}) \cdot (P_{PD} \cdot P_{OD} \cdot |\psi\rangle \otimes |\varphi_{OD}\rangle \otimes |\varphi_{PD}\rangle)$$

$$+ (I_N \otimes R_{OC} \otimes R_{PD}) \cdot (P_{PD} \cdot P_{OC} \cdot |\psi\rangle \otimes |\varphi_{OC}\rangle \otimes |\varphi_{PD}\rangle)$$

$$+ (I_N \otimes R_{OD} \otimes R_{PC}) \cdot (P_{PC} \cdot P_{OD} \cdot |\psi\rangle \otimes |\varphi_{OD}\rangle \otimes |\varphi_{PC}\rangle)$$

$$+ (I_N \otimes R_{OC} \otimes R_{PC}) \cdot (P_{PC} \cdot P_{OC} \cdot |\psi\rangle \otimes |\varphi_{OC}\rangle \otimes |\varphi_{PC}\rangle)\|^2$$

And because all 4 terms inside the bracket are orthogonal we obtain the sum of probabilities

$$p_T(PD) = \|P_{PD} \cdot P_{OD} \cdot |\psi\rangle\|^2 + \|P_{PD} \cdot P_{OC} \cdot |\psi\rangle\|^2$$

$$+ \|P_{PC} \cdot P_{OD} \cdot |\psi\rangle\|^2 + \|P_{PC} \cdot P_{OC} \cdot |\psi\rangle\|^2.$$

In this way, the measurement of the prediction eliminates any interference, and the probabilities are identical to classical probabilities.