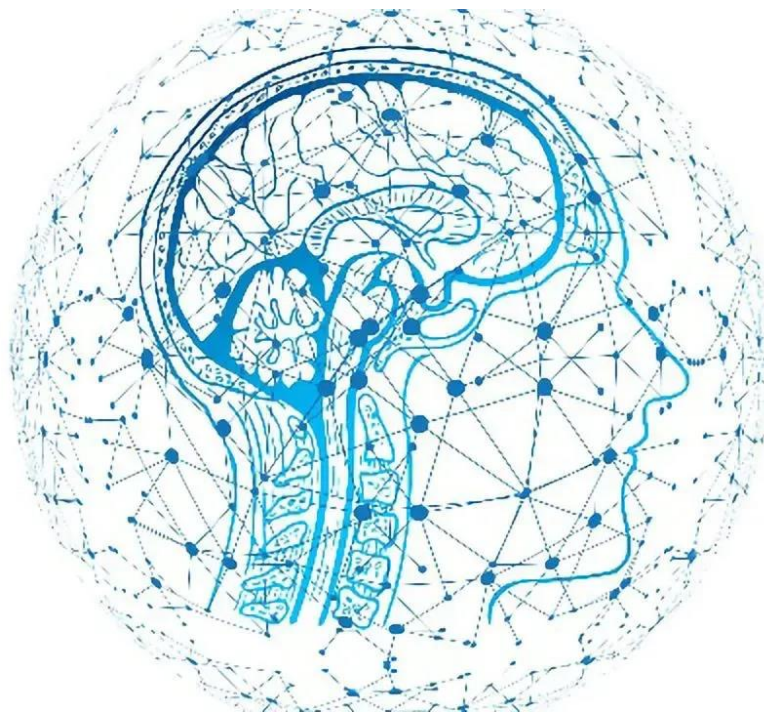


CONSCIOUSNESS VOL 5



Quantum Physics and the Existence of God

by

Stephen Priest

Faculty of Philosophy, University of Oxford, Oxford OX1 2JD, UK

Religions **2024**, *15*(1), 78; <https://doi.org/10.3390/rel15010078>

Submission received: 18 July 2023 / Revised: 16 December 2023 / Accepted: 5 January 2024 / Published: 9 January 2024

(This article belongs to the Special Issue [Minds as Creaturely and Divine](#))

Abstract

Interpretations of quantum physics are shown to presuppose the reality of consciousness. But if a minimal realism about the external world is true, then the consciousness presupposed by quantum reality cannot be only that of the scientific observer, cannot be only ‘local’ but must be ‘global’. Global consciousness is argued to have all and only the essential properties of God. Quantum reality depends on God’s consciousness and the physical world depends on quantum reality. Therefore, the physical world depends on God’s consciousness.

Keywords:

[God](#); [science](#); [religion](#); [quantum](#); [physics](#); [consciousness](#); [being](#); [presence](#)

1. Introduction

Interpretations of quantum physics presuppose the reality of consciousness. But if a minimal realism about the external world is true, then the consciousness presupposed by quantum reality cannot be only that of the scientific observer, cannot be only 'local' but must be 'global'. Global consciousness is argued to have all and only the essential properties of God. Quantum reality depends on God's consciousness and the physical world depends on quantum reality. Therefore, the physical world depends on God's consciousness.

We know, from the recent empirical confirmations of Bell's criticism of the Einstein–Podolsky–Rosen experiment (EPR)^{[1](#)}, that quantum reality does not depend on anything classical: there is no macroscopic Newton–Einstein world more fundamental than the quantum level. It follows that consciousness does not depend on, and is irreducible to, anything physical. If the quantum depends on consciousness, and consciousness were to depend on the physical, then the quantum would depend on the physical (via consciousness). But that is precisely ruled out by Bell's criticism of the EPR experiment and by subsequent practical work.^{[2](#)}

It has often been pointed out that, for reasons peculiar to it, the popular Copenhagen interpretation of quantum mechanics seems to presuppose consciousness. Here, it is argued that *no* interpretation escapes this presupposition. The reasons for this are not always so closely tied to the unique character of the interpretation in question. Nonetheless, interpreters of quantum mechanics are without exception forced to fall back upon consciousness. The presupposition is ontological, not just epistemological: consciousness is entailed by quantum reality, not just by inquiry into quantum reality, by what would make the theory true, not just by the consciousness of the theoretician.^{[3](#)}

It follows that some salient tenets of modernity are false: consciousness does not depend on and is not reducible to the physical world, in any sense of 'reducible', so consciousness is not a product of evolution. We should not be surprised by this result because, if consciousness were a product of evolution, consciousness would be an emergent property of the brain. But the brain, for all its anatomical complexity, is only billions of atoms in empty

space, and billions of atoms in empty space are neither logically nor causally sufficient for consciousness. Consciousness did not evolve.[4](#)

Some important theses are entailed: the Thomist theses that God is his existence, and God is actus purus, and the idealist thesis that the physical world depends on consciousness. Being, Presence, and Consciousness itself are, fundamentally understood, the persons of the Holy Trinity.

This paper is a summary of an unpublished book manuscript, and I do not pretend that there is not much more to be said about the problems and their putative solutions.

2. Consciousness and Quantum Physics

(1)

The Copenhagen Interpretation ('C')

In C, prima facie meaningful claims of the form: 'Particle Φ has momentum $\mathbf{p}$ but is not measured' or 'Particle Φ is at position $\mathbf{q}$ but is not measured' do not have truth values, do not express propositions, or are devoid of literal significance. This anti-realism is motivated by the impossibility of simultaneously measuring momentum and position, at a quantum level.

Observation, or the act of measurement itself, 'realises', makes the case, or makes actual, a definite momentum or a definite position, instantaneously. Of the possible momenta or positions consistent with the wave function, measuring actualises the possibility of Φ having $\mathbf{p}$ or actualises the possibility of Φ being at $\mathbf{q}$, but not both simultaneously.

Because measurement collapses the wave function, it is interaction with an observer outside the quantum system that effects the transition from indeterminacy to determinacy.

Although observation determines reality, a degree of realist explanation is retained by the notion of complementarity. For any two predicates whose application is mutually inconsistent:

'Complementarity' (*def.*) F and G are complimentary *iff*, at any time, t , if x is measured to be F , x cannot not be measured to be G , and if x is measured to be G , x cannot not be measured to be F , but x may be measured to be F at t , and measured to be G at t^1 .

For example, the accuracy with which momentum is measured, at a time, diminishes the accuracy with which position is measured and vice versa. One and the same reality may be described as a particle or as a wave, over time, with logical consistency.[5](#)

- *Does C Presuppose Consciousness?*

In C, observation collapses the wave function. But if there is observation then consciousness exists. Therefore, consciousness exists.

Why is it right that if there is observation then consciousness exists? There is a distinction between the following: $[\alpha]$ some cause brings about some effect, and either (classically) $[\beta]$ some cause brings about some [effect that is an] observation or (in a quasi-quantum idiom) $[\gamma]$ some [cause that is an] observation brings about some effect, because each of $[\beta]$ and $[\gamma]$ says more than $[\alpha]$. What is this extra information? What is it for some cause or some effect to be an observation? If there is observation:

[1]

Something is *something to* someone or other.

[2]

Something is *present to* someone or other.

[3]

Someone is *aware of something as* something or other.

These closely related relations only obtain if observation is essentially conscious observation. To see that [1] to [3] are entailments of consciousness, consider the necessary falsity of their converse. There is consciousness and: *Nothing is anything to anyone. Nothing is present to anyone. No one is aware of anything as anything.*

Admittedly, proponents of C often allow a nonconscious item, such as a Geiger counter, a clock, a photographic plate, or some other object that is not conscious as ‘the observer’. However, a Geiger counter, a clock, or a photographic plate is itself a physical system that admits of a quantum description. In C, that description will entail that the state of the physical ‘observer’ is itself indeterminate until it is observed in its turn. It might be tempting to introduce *another* nonconscious physical ‘observer’ to bring about the determinate state of the Geiger counter or clock—an eye or a brain perhaps. But to do so reintroduces the same problem, for an eye or a brain is also a physical system that admits of a quantum description. For this reason, C makes sense only if wave-function collapse is in every case ultimately caused by a conscious nonphysical observer. Indeed, Bell’s criticism of EPR rules out any classical or macroscopic physical reality making possible anything quantum.

In addition, it would be vacuously circular to argue that the cause of wave-function collapse is only either classical or quantum: classical physics depends on quantum physics but not vice versa, and any quantum event that causes any quantum event itself ultimately stands in need of non-quantum explanation. We need to know: *why is there a quantum world?*

Any measurement entails some conscious act of observation, because not only a physical event is sufficient for wave-function collapse. Anything is only an observation if there is conscious observation. Any putative chain of non-conscious observations has to be arrested in some conscious observation that makes non-conscious observation observation. If there is observation then there is consciousness.

• **Argument that C presupposes consciousness:**

$$(\psi \rightarrow \Phi i) \tag{1}$$

$$(\psi \rightarrow \Phi i) \rightarrow \exists(x) (\Theta x) \tag{2}$$

$$\exists(x) (\Theta x) \rightarrow \exists(x) (\odot x) \tag{3}$$

$$\exists(x) (\odot x) \tag{4}$$

(2)

The Many Worlds Interpretation (MW)

In MW, there is no quantum indeterminacy because any quantum event splits the universe into two numerically distinct, wholly determinate, universes, one in which a particle has position, the other in which it has momentum (or one in which Schrödinger's cat is dead and one in which Schrödinger's cat is alive). The many universes are proliferations of 'the' universe. There is no wave-function collapse, because there is wave-function branching. Quantum physics applies to the set of universes as a whole, or multiverse, and not to any member of the set, or individual universe. All numerically distinct universes exist in one and the same space–time. In MW, all theoretical, logical, and mathematical possibilities are realised.[6](#)

The universal wave function obeys the same deterministic, reversible laws at all times. The theory preserves wave–particle duality because nothing has a definite momentum and position simultaneously. In some versions of MW there are no particles but there are waves, because a necessary condition for being a particle is possessing definite momentum and position simultaneously.

• *Does MW presuppose consciousness?*

Proponents of MW sometimes regard it as a merit of the interpretation to eschew the reality of consciousness, at least as seemingly implied by collapse of the wave function in the Copenhagen interpretation. But can consciousness be eschewed?

In MW, any quantum event splits the universe into two numerically distinct universes. But what is it for some universe to be ‘the’ universe? ‘The’ universe is ‘our’ universe, the universe that you and I inhabit, and so, in a clear sense, consciousness is presupposed, after all. De Witt (who coined the term ‘Many Worlds’) says:

[...] every quantum transition taking place on every star, in every galaxy, in every remote corner of the universe is splitting our local world on earth into myriads of copies of itself.

([De Witt 1970, p. 33](#))

Some universe being ‘this’ or ‘the’ universe presupposes at least one ‘I’ user, but there is only an ‘I’ user if there is self-consciousness and there is only self-consciousness if there is consciousness. Therefore, MW presupposes consciousness.

Also, in MW, there is observation in some universe, which does not extend to any distinct universe. Therefore, in MW, there is observation. But if there is observation, there is consciousness. Therefore, wittingly or not, in MW, there is consciousness.

In MW, universes are therefore epistemologically distinguished by the impossibility of observation (or transference of information) between worlds. In the absence of any other criterion, in MW, the possibility or impossibility of observation (or transference of information) is the sole criterion for individuating worlds. Because observation or information presuppose consciousness, as MW stands, consciousness has an essential role in constituting universes as distinct.

• **Argument that MW entails the existence of consciousness:**

$$\exists(x) (\Theta x, w^1) \ \& \ \forall(w) \neg(w = w^1) \rightarrow \neg(\Theta x w) \tag{1}$$

$$\exists(x) (\Theta x) \tag{2}$$

$$\exists(x) (\Theta x) \tag{3}$$

$$\exists(x) (\Theta x) \rightarrow \exists(x) (\odot x) \tag{4}$$

$$\exists(x) (\odot x)$$

• The Consistent Histories Interpretation (CH) ([Griffiths 1984](#))

In CH, there exists a consistent set of possible histories of any quantum system, where a history is a sequence of propositions indexed to times. Probabilities are assigned to the disjunct possible histories. The allocation of probabilities to propositions in a history is only within a theory, or belief system, or web of belief.

There is no wave-function collapse, so no dependence of wave-function collapse on observation or measurement (as exists in C).

- *Does CH entail the existence of consciousness?*

In CH, the allocation of probabilities to propositions in a history is only within a theory, or belief system, or web of belief. If there is a theory, or a belief system, or a web of belief, then there is consciousness. Therefore, there is consciousness.

If there are theories, beliefs, etc., then there is consciousness because there can only be theories, beliefs, etc. if there are theoreticians, believers, etc. and there are only theoreticians, believers, etc. if there is at least one conscious theoretician, believer, etc. For a reductio ad absurdum entertain: 'There are beliefs but no one has ever believed (disbelieved) anything', 'There are theories never held by anyone', etc.

In platonic realism, indeed, there are mind-independent theories, consciousness-independent beliefs, etc. But this is at most plausible only for beliefs only in the sense of the propositional content of beliefs, theories in the sense of the propositional content of theories. Beliefs in the sense of propositional attitudes, theories as they exist in science, presuppose conscious believers, scientists, for their being and for their being what they are.

- **Argument from belief, that CH entails the existence of consciousness:**

$$\exists(x) Bxp \tag{1}$$

$$\tag{2}$$

$$\exists(x) Bxp \rightarrow \exists(x) (\odot x) \tag{3}$$

$$\exists(x) (\odot x)$$

In CH, at least one history exists. If a history exists, then there is past, present, and future. If there is past, present, and future, then consciousness exists. Therefore, consciousness exists.

It is right that if there is past, present, and future then consciousness exists because any distinction between past, present, and future is a distinction between *someone's* past, present, and future, and being someone entails being conscious.⁷ Anyone is only anyone if they are conscious. Any distinction between past, present, and future therefore entails the existence of consciousness.⁸ The presupposition is ontological not just epistemological: A history is a history of quantum reality not just a history of the inquirer. Therefore, in CH, quantum reality presupposes consciousness.

- **Argument, from history, that CH entails the existence of consciousness:**

$$\exists(x) Hx \tag{1}$$

$$\exists(x) Hx \rightarrow \exists(x) (\Theta x) \tag{2}$$

$$\exists(\Theta x) \rightarrow \exists(x) (\odot x) \tag{3}$$

$$\exists(x) (\odot x) \tag{4}$$

In CH, one is left wondering what it consists of for one probable history rather than another to be realised, for one of a range of possibilities to be actual or actualised. Copenhagen measurement is evaded only by leaving the truth values of the propositions of quantum mechanics at the level of probability. But once we admit actuality, we admit consciousness.

(3)

The Ensemble Interpretation (E)

In E, quantum descriptions do not apply to any individual quantum phenomenon, but only to a large number, or cloud, or 'ensemble', of similar quantum phenomena.⁹ Quantum descriptions are only probabilistic and quantum reality is only an abstract statistical quantity that has no ontological status outside mathematics.

Of course, all interpretations use and hence presuppose the existence of mathematics. But, in the case of E, the presupposition is salient because the existence of the quantum world itself is allegedly only mathematical.

- *Does E presuppose consciousness?*

In E, quantum reality has only mathematical ontological status. If intuitionism is correct, the existence of mathematics depends on the existence of consciousness. Intuitionism is correct. Therefore, consciousness exists.

Is there any reason to accept intuitionism? Intuitionism is idealism about mathematics: there is no mathematical truth that outstrips proof. But proving is a conscious activity. So, if there is mathematical truth, there is consciousness. There is no mathematics unless something can be counted. But counting is a conscious activity. So, there is no mathematics without consciousness. Mathematics is constructed not discovered. If there is no construction unless there is conscious construction, again, consciousness is presupposed.

In a sense, intuitionism is an ontologically cautious theory. In interpreting mathematics, we postulate nothing beyond what we strictly know. Intuitionism's realist repudiation arguably violates Ockham's Razor.[10](#)

If intuitionism (or similar) is false, then mathematics is 'out there'. There is mathematical truth irrespective of the possibility of proof. There are numbers even if there is or could be no counting. Mathematics is discovered, not invented. However, we are still left with: *What is the ontological status of mathematics? Why is there any mathematics? Why is the universe mathematical?* We might well need consciousness anyway, and this time on a global or cosmic scale, to begin to answer these questions.

• **Intuitionist Argument that E Presupposes Consciousness:**

$$\exists(x) (x = n) \tag{1}$$

$$\exists(x) (x = n) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

If mathematics presupposes consciousness, and if the physical world presupposes mathematics, then the physical world presupposes consciousness. The presupposition is ontological.

(4)

The Branching Space-Time Interpretation

BST entails the essentials of MW and may be rightly thought a variant of it. In MW, it is quantum reality which branches within one and the same space-time. In BST, it is space-time itself which branches into numerically distinct space-times. In the cat analogy, space-time branches into two space-times, one in which Schrödinger's cat is dead and one in which Schrödinger's cat is alive. In the terminology of histories, branching into histories pertains to the topology of sets of events standing in causal relationships within space-time.

BST faces the same sorts of logical problems as MW but about space-times rather than worlds.

• *Does BST Presuppose Consciousness?*

BST presupposes MW and MW presupposes consciousness. Therefore, BST presupposes consciousness.

The argument is valid because of the transitivity of 'entails': if p entails q and q entails r , then p entails r . Anything entails all and only the entailments of what it entails.

The argument is not sound if the first premise is false, for example because the two interpretations differ in entailing different numbers of

space–times. We could revise the first premise to read: ‘BST presupposes the essentials of MW’ or ‘the part of MW that entails the existence of consciousness’, or similar. Then, the argument is not only valid but sound.

• **Argument that BST presupposes consciousness:**

$$\exists(x) (\Theta x, s - t^1) \ \& \ \forall(s - t) - (s - t = s - t^1) \rightarrow -(\Theta x s - t) \quad (1)$$

$$\exists(x) (\Theta x) \quad (2)$$

$$\exists(x) (\Theta x) \rightarrow \exists(x) (\odot x) \quad (3)$$

$$\exists(x) (\odot x) \quad (4)$$

It is the branching of space–time itself which presupposes consciousness, not just any theory or knowledge of this branching. It is therefore quantum reality which presupposes consciousness in BST, not just knowledge of quantum reality.

(5)

The de Broglie–Bohm Interpretation (B-B)

In B-B, any particle has a definite position and a definite momentum. Position and momentum are hidden variables because the investigator cannot know them both simultaneously and precisely, because of Heisenberg’s uncertainty principle.

The wave function depicts the configuration of the universe as a whole. Causation is non-local, and so satisfies Bell’s inequality and is deterministic. The momentum of a particular particle depends on the value of the guiding equation, which depends on the configuration of the universe as a whole, minus that particle, that is, the position of every other particle. Particles are guided by the ‘pilot wave’: $\psi(q, t) \in \mathbf{C}$. The wave function does not collapse, because collapse is only phenomenological collapse, not ‘real’ or ‘objective’ collapse.

The measurement problem is putatively solved because, at any given time, any particle has definite position and momentum. B-B only entails the existence of one space–time and one world, so does not suffer some of the logical drawbacks of MW, CH, and BST.

The strategy is to save both classical determinism and realism: Causes necessitate their effects, and events either happen or do not, irrespective of the existence or nature of their observer. Uncertainty is only epistemological, not ontological. If B-B is correct, two of the apparent consequences of quantum physics which appalled Einstein are avoided: God does not play dice, and the physics is not subjective but ‘out there’. B-B is a hidden variable

theory. In a plausible construal, Bohm's interpretation entails that, if anything exists, it exists now (in a classical sense).[11](#)

- *Does B-B Presuppose Consciousness?*

In B-B, wave function collapse is only phenomenological wave function collapse. But, if there is any phenomenology, then consciousness exists. There is phenomenology. Therefore, consciousness exists.

- **Argument that B-B presupposes consciousness:**

$$Bx [(|\psi\rangle \rightarrow |\Phi_i\rangle)] \ \& \ -(|\psi\rangle \rightarrow |\Phi_i\rangle) \tag{1}$$

$$Bx [(|\psi\rangle \rightarrow |\Phi_i\rangle)] \tag{2}$$

$$Bx [(|\psi\rangle \rightarrow |\Phi_i\rangle)] \rightarrow \exists(x) (\Theta x) \tag{3}$$

$$\exists(x) (\Theta x) \tag{4}$$

$$\exists(x) (\Theta x) \rightarrow \exists(x) (\odot x) \tag{5}$$

$$\exists(x) (\odot x)$$

If there is any phenomenology, then consciousness exists, because there is no phenomenology unless something is something *to* someone and so, in an obvious sense, present to consciousness.

Of course, this sort of argument applies to *any* theory that grants the existence of phenomenology. It is characteristic of B-B to make phenomenology conspicuously essential to quantum physics but the presupposition is what we are calling 'epistemological', i.e., presupposed by the epistemology, rather than 'ontological', i.e., presupposed by quantum reality itself.

- **Argument that Bohm's interpretation presupposes consciousness:**

$$\exists(x) (\Phi x) \ \& \ \exists(y) (Dy) \ \& \ Rxy \tag{1}$$

$$\exists(x) (\Phi x) \ \& \ \exists(y) (Dy) \ \& \ Rxy \tag{2}$$

$$\exists(x) (\Phi x) \ \& \ \exists(y) (Dy) \ \& \ Rxy \rightarrow \exists(z) (\odot z) \tag{3}$$

$$\exists(z) (\odot z)$$

Suppose Bohm's interpretation entails that, if anything is true, it is true *now*. A point of time classically conceived is not sufficient for that time's being now, for its being 'the time it is'. What else is necessary?

If anything exists, it exists now. Something exists. Something exists now. 'Now' means 'when I am' so a time's being now is its being simultaneous with when I am or when some 'I' user is. Being an 'I' user entails being

conscious. Any time's being now entails the existence of consciousness. Therefore, consciousness exists.

We may imagine robotic simulations of I-use but, in the absence of consciousness, they remain simulations.

(6)

Relativistic Quantum Physics (RQP)

In RQP, a system might be in a single, 'collapsed', eigenstate to an observer at a time but in a superposition of two or more states to a different observer at that time.

'State' refers to nothing intrinsic to the observed system itself, but denotes the relation, or correlation, between that system and its observers. States of *all* objects and events are really relations between observer and observed (irrespective of whether they are Newtonian, quantum, conscious, etc.). The state vector of quantum physics is construed as a description of the correlation between the observed system and some degree of freedom in the observer.

In RQP, a detection event establishes a relation between the quantised field and the detector, thus avoiding the wave–particle ambiguity in Heisenberg's uncertainty principle. Any measurement event is an ordinary physical interaction, the establishing of the correlation between measuring device and measured.

RQP exploits the relativity of simultaneity in the special theory of relativity: the same events might appear temporally ordered in a different way to different observers: e might appear simultaneous with e_1 to A but earlier than e_1 to B and later than e_1 to C.

- *Does RQP presuppose consciousness?*

In RPQ, if there exists a particle with at least one property then there exists an observer of that particle, thereby necessarily related to that particle. There exists a particle with at least one property. Therefore, there exists an observer. If there exists an observer, then consciousness exists. Therefore, consciousness exists.

For the soundness of the argument, we rely here on $(\exists(x) (\Theta x) \rightarrow \exists(x) (\odot x))$ as an assumption.

- **Argument that Relativistic Quantum Physics presupposes consciousness (I):**

$$\forall(x) ((\Phi x \ \& \ Fx) \rightarrow \exists(y)(\Theta y) \ \& \ (Rxy)) \tag{1}$$

$$\exists(x) ((\Phi x \ \& \ Fx) \rightarrow \exists(y)(\Theta y) \ \& \ (Rxy)) \tag{2}$$

$$\exists(x) ((\Phi x \ \& \ Fx) \rightarrow \exists(y)(\Theta y) \ \& \ (Rxy)) \tag{3}$$

$$\exists(y)(\Theta x) \quad (4)$$

$$\exists(y) (\Theta x) \rightarrow \exists(x) (\odot x) \quad (5)$$

$$\exists(x) (\odot x)$$

In RQP, there exists a particle and a detector of that particle. If a detector exists, then consciousness exists. Therefore, consciousness exists.

Is it right that, if a detector exists, then consciousness exists? Clearly, there may exist a physical system internally related such that one part receives a pattern of causal impacts from another, for example, a distribution of particles fired from a source, hitting a photographic plate. However, it is only right to call the photographic plate 'a detector' if there is awareness of it and the pattern of causal impacts it receives (either by it or by something else). (a) 'D detects a pattern of causal impacts' says more than: (b) 'D is subject to a pattern of causal impacts'. (a) entails (b) but (b) does not entail (a). (We envisage the impacts in (a) as *on* D.) If (b) is insufficient for (a), then what else is necessary? 'There is consciousness that D is subject to a pattern of causal impacts' is sufficient for (a).

• **Argument that Relativistic Quantum Physics presupposes consciousness (II):**

$$\forall(p) p \rightarrow pt^0 \quad (1)$$

$$\quad (2)$$

$$\exists(p) p \rightarrow pt^0 \quad (3)$$

$$\exists(p) p \rightarrow pt^0 \rightarrow \exists(y)(\Theta x) \quad (4)$$

$$\exists(y) (\Theta x) \rightarrow \exists(x) (\odot x) \quad (5)$$

$$\exists(x) (\odot x)$$

The presupposition of consciousness in RQP is ontological, not just epistemological, because in RQP quantum reality is essentially related to an observer but the existence of an observer presupposes consciousness (at least somewhere along the line). Therefore, quantum reality itself presupposes consciousness.

(7)

The Quantum Information Interpretation (ontological) (QI(O))

In J. A. Wheeler's 'it from bit' ontology, a quantum state is a mental construct of the observer and quantum realities are consequences of information.

- *Does QI(O) presuppose consciousness?*

In QI(O), a quantum state is a mental construct of the observer. But if there is an observer there is consciousness. Therefore, consciousness exists. More directly: if there are mental constructs, then consciousness exists. There are mental constructs, so consciousness exists.

Common sense and classical physics are committed to the mind-independent existence of the physical world. However, QI(O) calls this independence into question. If QI(O) is correct, we need a theory of the observer. Berkeleyan idealism, or something close to it, is then a plausible quantum philosophy.

- **Argument that QI(O) presupposes consciousness (I):**

$$\exists(x) (\Theta x) \tag{1}$$

$$\exists(x) (\Theta x) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

In QI(O), quantum realities are consequences of information. If information exists then consciousness exists. Therefore, consciousness exists.

'Consequence' is ambiguous between 'causal consequence' and 'logical consequence'. Construed causally, quantum reality is an emergent property of information. Information, in some way, is causally sufficient for quantum phenomena. Construed logically, quantum reality is essentially constituted of information or analytically reducible to information. The first reading yields a Cartesian dualism, the second a Berkeleyan immaterialism. In either reading, consciousness is straightforwardly presupposed, so long as the existence of information entails the existence of consciousness.

But: What is information? What is it for some cause or some effect to be information? Information is intellectual or perceptual content which informs.

[1]

Information is *information to* someone or other.

[2]

Information is *present to* someone or other.

[3]

Someone is *aware of information as* something or other.

These closely related relations may only obtain if information is essentially information for consciousness. If you are acquainted with information, it is something to you, rather than just something there is. It is present to you, both in the sense of being 'here' and in the sense of being 'now'. Somewhere being here, and some time being now, directly presuppose consciousness. 'Here' means 'where I am' and 'now' means 'when I am', where being an 'I' user entails being a conscious being.

There is information without consciousness only if chains of information are arrested in information present to consciousness. Why is this right? There is a distinction between the following:

[α] Some cause brings about some effect.

$$\exists(x) \exists(y) (Cxy)$$

and either (classically):

[β] Some cause brings about some [effect that is] information.

$$\exists(x) \exists(y) (Cxy) \& (i y)$$

or (in a quasi-quantum idiom):

[γ] Some [cause that is] information brings about some effect.

$$\exists(x) \exists(y) (Cxy) \& (i x)$$

If there is information, then consciousness exists, because there is information without consciousness only if chains of information are arrested in information present to consciousness. If there is information, some consciousness is *informed*.

• **Argument that QI(O) presupposes consciousness (II):**

$$\exists(x) \exists(y) (Cxy) \& (i x) \tag{1}$$

$$\exists(x) (i x) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

Quantum reality presupposes information but there is no information unless consciousness exists, so quantum reality itself presupposes consciousness. The presupposition is ontological, not only epistemological.

(8)

The Quantum Information Interpretation (Epistemological) (QI(E))

In QI(E), quantum mechanics is true of an observer's knowledge of the world, not the world itself, because there is no such thing as the collapse of the wave function over and above an observer acquiring information from the measurement of some event.[12](#) The 'reduction of the wave-packet' takes

place in the consciousness of the observer, not because of any physical process which takes place there, but only because the state is a construct of the observer. It is not an objective property of the physical system.

A state is information, obtained from knowledge of how a system was prepared, used for predicting future measurements. A state is not a state of anything objective in the sense of 'observer independent'. A quantum mechanical state is a summary of the observer's information about an individual physical system's changes by dynamical laws whenever the observer acquires new information about the system through the process of measurement. The state vector is not an objective property of the system.

- *Does QI(E) presuppose consciousness?*

Quantum physics is nothing over and above, or reducible to, the knowledge of quantum physics. If there is knowledge then consciousness exists. There is knowledge. Therefore, consciousness exists. The existence of certain quantum mechanical facts is the existence of certain items of knowledge. So long as knowledge entails the existence of consciousness, consciousness is clearly presupposed.

- **Argument that QI(E) presupposes consciousness:**

$$\exists(p) (Kxp) \tag{1}$$

$$\exists(p) (Kxp) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

In a sense, the presupposition of consciousness in QI(E) is epistemological. But this is because, in QI(E), quantum reality itself is only epistemological: it does not exceed the knowledge of it. Quantum reality is not 'out there'. But, in so far as there is quantum reality, consciousness is presupposed.

(9)

The von Neumann/Wigner Interpretation (N/W)

In N/W, consciousness collapses quantum reality and, in von Neumann's version, the entire physical universe is subject to the Schrödinger equation (the universal wave function). In N/W, physics needs consciousness in order to be complete.

- *Does N/W entail the existence of consciousness?*

N/W overtly implies that quantum physics presupposes consciousness, because consciousness collapses the wave function. There is wave-function collapse and, if there is wave-function collapse, then consciousness exists. Therefore, consciousness exists.

- **Argument that N/W presupposes consciousness:**

$$[(|\psi\rangle \rightarrow |\Phi_i\rangle)] \quad (1)$$

$$[(|\psi\rangle \rightarrow |\Phi_i\rangle)] \rightarrow \exists(x) (\odot x) \quad (2)$$

$$\exists(x) (\odot x) \quad (3)$$

In N/W, the entire physical universe is subject to the Schrödinger equation (the universal wave function) and so dependent on its holding. But the Schrödinger equation holding depends on consciousness. Therefore, the entire physical universe depends upon consciousness. The inference is valid because of the transitivity of 'depends on': if a depends on b and b depends on c , then a depends on c .

- **Argument that the entire physical universe depends upon conscious observation:**

$$\exists(x) (x = w) \rightarrow (|\psi\rangle \rightarrow |\Phi_i\rangle) \quad (1)$$

$$(|\psi\rangle \rightarrow |\Phi_i\rangle) \rightarrow \exists(x) (\odot x) \quad (2)$$

$$\exists(x) (\odot x) \quad (3)$$

$$\exists(x) (x = w) \quad (4)$$

Obviously, if the entire physical universe depends on conscious observation, then consciousness is presupposed ontologically, not just epistemologically, in our senses of these terms.

(10)

The Wheeler/Stapp Interpretation (W/S)

In W/S, consciousness is essential to bringing the universe into existence and, although the universe began, consciousness has always existed.

- *Does W/S entail the existence of Consciousness?*

In W/S, consciousness is essential to bringing the universe into existence. The universe came into existence. So, in W/S, if there is a universe, then consciousness exists and a universe exists. It follows that consciousness exists.

We do not know with certainty that the universe did come into existence, even though at the time of writing most physicists think the universe began with the Big Bang. It is hard to rule out a priori the existence of a universe

infinite in backwards time. Plus, even if this universe began, it is hard to rule out a priori a chain of predecessor universes.

Is the final inference valid? Logically and theoretically, consciousness could bring the universe into existence but then itself cease to exist. There seems no incoherence in such a supposition. On the other hand, if consciousness is necessary to sustain the universe, then consciousness has to exist at least so long as there are quantum events. If consciousness is necessary for time (say, because quantum reality is necessary for time), the assumption that there was a time when there was no consciousness must be false. At any time, 'There is consciousness' is true (even if consciousness is intrinsically atemporal). Either way, consciousness exists at least *sub specie aeternitatis* in W/S.

• **Argument that W/S entails consciousness:**

$$\exists(x) (x = w) \rightarrow \exists(x) (\odot x) \quad (1)$$

$$\exists(x) (x = w) \quad (2)$$

$$\exists(x) (\odot x) \quad (3)$$

Because, in W/S, consciousness is essential to bringing the universe into existence, consciousness is presupposed ontologically not just epistemologically.

(11)

The Many-Minds Interpretation (MM)

In MM, the mind of the observer splits into distinct minds, none of which is (numerically identical with) the mind that splits (one, say, thinking that Schrödinger's cat is dead, the other thinking that Schrödinger's cat is alive). A mind that results from a split occupies a different universe from any other mind that results from the same split.

MM entails a version of MW. But what it consists of for such universes to be distinct is those minds being distinct. In MW, the universe splits into distinct universes, none of which is numerically identical with the universe that splits. In MM, a mind splits into distinct minds, none of which is numerically identical with the mind that splits.

• *Does MM Presuppose Consciousness?*

In MM, there exists a mind. But, if there are any minds, then consciousness exists. Therefore, consciousness exists.

One might object to 'If there are any minds, then consciousness exists' that there is artificial intelligence: computation which in many respects outstrips human achievements, exhibited by non-conscious systems. If a

capacity to solve complex computational problems, speedily and accurately, is sufficient for the existence of a mind, then computers are or have minds.

On the other hand, a computer is not aware of anything whatsoever. We may simulate self-awareness by programming it with information about itself, equipping it with self-scanning devices, etc. but it remains without one iota of consciousness. There is a drastic sense in which any computer does not know what it is doing. It has not the slightest *awareness* of what it is doing. The many-minds interpretation requires minds that are aware, and minds, so understood, entail the existence of consciousness. Some people argue that (some) computers are or could be conscious. But, if they are right, the existence of conscious computers does not undermine but lends support to the premiss: 'If there are any minds, then consciousness exists'.

- **Argument that MM presupposes consciousness:**

$$\exists(x) (Mx) \tag{1}$$

$$\exists(x) (Mx) \rightarrow \exists(y) (\odot y) \tag{2}$$

$$\exists(y) (\odot y) \tag{3}$$

Like MW above, MM presupposes consciousness ontologically, not just epistemologically. Quantum reality, not just knowledge of quantum reality, presupposes consciousness.

(12)

The Relational Interpretation (R)

R is a version of M. In R, quantum mechanical descriptions are generated from different perspectives, are equally objective, and correspond to physical reality. The physical world is essentially or fundamentally relational. In particular, relations do not depend on anything non-relational.[13](#)

- *Does R presuppose consciousness?*

Any quantum description is derived from a 'perspective'.[14](#) A perspective presupposes consciousness. Therefore, consciousness exists.[15](#)

- **Argument that R presupposes consciousness (I):**

$$\forall(p) Kxp \rightarrow \exists(x) (\text{perspective } x) \tag{1}$$

$$\exists(x) (\text{perspective } x) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

Any quantum description is derived from a ‘perspective’. But any perspective is someone’s perspective.[16](#) A perspective essentially belongs to an observer and an observer is essentially conscious. Therefore, consciousness exists. Rovelli himself recognises the role of the observer in our making sense of quantum physics.[17](#)

• **Argument that R presupposes consciousness (II):**

$$\forall(p) Kxp \rightarrow \exists(x) (\text{perspective } x) \quad (1)$$

$$\forall(p) Kxp \rightarrow \exists(x) (\text{perspective } x) \quad (2)$$

$$\exists(x) (\text{perspective } x) \rightarrow \exists(x) (\Theta x) \quad (3)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (4)$$

$$\exists(x) (\odot x)$$

Any quantum description depends on a measurement.[18](#) But any measurement essentially presupposes an observer. An observer is essentially conscious. Therefore, consciousness exists.

• **Argument that R presupposes consciousness (III):**

$$\forall(p) Kxp \rightarrow \exists(x) (\text{measurement } x) \quad (1)$$

$$\forall(p) Kxp \rightarrow \exists(x) (\text{measurement } x) \quad (2)$$

$$\exists(x) (\text{measurement } x) \rightarrow \exists(x) (\Theta x) \quad (3)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (4)$$

$$\exists(x) (\odot x)$$

Indeed, Rovelli himself takes the view that the results of measurement only have meaning in relation to an observing system or observer.[19](#)

Any quantum description is derived from ‘actuality’.[20](#) Actuality presupposes consciousness. Therefore, consciousness exists.

• **Argument that R presupposes consciousness (IV):**

$$\exists(x) (\text{actuality } x) \quad (1)$$

$$\exists(x) (\text{actuality } x) \rightarrow \exists(x) (\Theta x) \quad (2)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (3)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (4)$$

$$\exists(x) (\odot x)$$

Any quantum description entails a transference of information. Information presupposes consciousness. Therefore, there is consciousness.

Consciousness is someone's consciousness, say 'our' consciousness. But: who are 'we'?²¹ Could there not, after all, be information but no observer?²² But even in a Humean or Buddhist 'no-self' theory that there is no self, subject, soul, conscious centre, or 'owner' of experience, there is experience, so consciousness. The existence of consciousness is required for mere events to be experiences and for mere processes to be information.

• **Argument that R presupposes consciousness (V):**

$$\exists(x) (ix) \tag{1}$$

$$\exists(x) (ix) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

In R, quantum reality is essentially related to consciousness, so consciousness is presupposed ontologically.

(13)

The Time-Symmetric Interpretation (TS)

In TS, there is backwards causation: future events bring about past events, as past events bring about future events. A single measurement cannot fully determine the state of a system (so the theory is a hidden variable theory). But, given two measurements performed at different times, it is possible to calculate the exact state of the system at all intermediate times. The collapse of the wave function is therefore not a physical change to the system, just a change in our knowledge of it due to the second measurement. Similarly, entanglement is not an objective physical state but an illusion created by ignoring backwards causation. The point where two particles appear to 'become entangled' is just a point where each particle is being influenced by events that occur to the other particle in the future. Modification of the equations of quantum mechanics are symmetric with respect to time reversal.

• *Does TS presuppose consciousness?*

In Ts, future events bring about past events, as past events bring about future events. So, in Ts, there is history. If there is history, then there is consciousness. Therefore, there is consciousness.

• **Argument that TS presupposes consciousness (I):**

$$\exists(x) (Hx) \tag{1}$$

$$\exists(x) (Hx) \rightarrow \exists(x) (\Theta x) \quad (2)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (3)$$

$$\exists(x) (\odot x) \quad (4)$$

The collapse of the wave function is just a change in our knowledge of it. So, in TS, there is knowledge. If there is knowledge, then there is consciousness. Therefore, there is consciousness.

• **Argument that TS presupposes consciousness (II):**

$$\exists(x) (Kxp) \quad (1)$$

$$\exists(x) (Kxp) \rightarrow \exists(x) (\Theta x) \quad (2)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (3)$$

$$\exists(x) (\odot x) \quad (4)$$

In TS, entanglement is an illusion. So, in TS there is illusion. If there is illusion, then there is consciousness. Therefore, there is consciousness.

• **Argument that TS presupposes consciousness (III):**

$$\exists(x) (ix) \quad (1)$$

$$\exists(x) (ix) \rightarrow \exists(x) (\Theta x) \quad (2)$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \quad (3)$$

$$\exists(x) (\odot x) \quad (4)$$

In TS, consciousness is presupposed ontologically, because history is the history of quantum reality and history presupposes consciousness, so quantum reality presupposes consciousness.

(14)

The Quantum Bayesian Interpretation (QB)

In QB, quantum reality is a measure of the observer's degree of belief about or knowledge of the behaviour of the observed system. [23](#)

• *Does QB presuppose consciousness?*

Quantum reality is a measure of the observer's degree of belief about the system.

There is belief. If there is belief, then there is consciousness. Therefore, there is consciousness.

• **Argument that QB presupposes consciousness (I):**

$$\exists(x) (Bxp) \tag{1}$$

$$\exists(x) (Bxp) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

In QB, quantum reality is a measure of the observer's knowledge of the behaviour of the system. So, there is knowledge. If there is knowledge, then there is consciousness. Therefore, there is consciousness.

• **Argument that QB presupposes consciousness (II):**

$$\exists(x) (Kxp) \tag{1}$$

$$\exists(x) (Kxp) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

Some systems are observed systems. If anything is observed, then there is consciousness. There is consciousness.

• **Argument that QB presupposes consciousness (III):**

$$\text{Some systems are observed systems.} \tag{1}$$

$$\text{If anything is observed, then there is consciousness.} \tag{2}$$

$$\text{There is consciousness.} \tag{3}$$

$$\exists(x) (\Theta x) \tag{1}$$

$$(\Theta x) \rightarrow \exists(x) (\odot x) \tag{2}$$

$$\exists(x) (\odot x) \tag{3}$$

In QB, the presupposition of consciousness is in a sense epistemological: quantum reality does not exceed the measure of the observer's degree of belief about or knowledge of the behaviour of the observed system, and allocations of probability are subjective (that is, are

only degrees of belief by the inquiring subject). Yet, to the extent to which quantum reality exists in QB, consciousness is presupposed.

The existence of the scientist, in the most intimate and direct sense, implies the existence of consciousness and quantum physics is fundamental science. So, it follows that there is no science without the scientist so no science without consciousness. Throughout history, there could not have been the investigations we call 'science' without scientific investigators. Science does not have the *method* it has without consciousness.

But quantum physics forces upon us a new and far more radical conclusion: science does not have *the subject matter* it has without the scientist. Consciousness *actualises*.[24](#)

Bell's criticism of the EPR experiment showed that there is no deterministic, macroscopic physical layer of reality which makes the quantum world possible. There is, therefore, nothing physical for the consciousness presupposed by the quantum world to depend on or be reducible to. Suppose 'prior' and 'fundamental' are understood like this:

[PR] 'prior' (*def.*) *a* is prior to *b* iff if not *a* then not *b*;

[F] 'more fundamental than' (*def.*) *a* is more fundamental than *b* iff if not *a* then not *b* but not if not *b* then not *a*;

[AF] 'absolutely fundamental' (*def.*) *a* is absolutely fundamental iff if not *a* then there is not anything but nothing except *a* is necessary for *a*.

It remains to be seen whether consciousness is absolutely fundamental. But consciousness is not just prior to, but ontologically more fundamental than, anything physical.

Given these conclusions, there is a clear need for science to admit a new category, even more fundamental than energy or matter: *consciousness*. A Theory of Everything, to be worthy of the name, will have to entail a theory of consciousness.

3. The Existence of God

Call the consciousness that exists if the quantum world exists 'quantum consciousness'. Quantum consciousness makes the universe possible. In traditional monotheism, God makes the universe possible. So, is quantum consciousness God? There is no wholly uncontroversial definition of 'God' but suppose God is this:

'God' (*def.*) *x* is God iff *x* is the unique, infinitely powerful, infinitely knowing, infinitely benevolent, necessary, omnipresent, atemporal ground of

everything, a person, and the Creator and sustainer of the universe: pure actuality.

Quantum consciousness *ontologically* inherits those properties entailed by being necessary for quantum reality.²⁵ Quantum consciousness, like anything, also presupposes whatever is necessary for there being anything whatsoever but in a special way: as properties of itself. So far as the finite mind may discern, the conjunction of these two sets of properties is all and only the essential properties of God, as understood in the great monotheistic religions. The existence of the universe entails, if not exactly God, then something so close to God as to be indistinguishable from God.

For brevity, I write ‘quantum’ and ‘universe’. But physics is open to the existence of more than one universe and regards the classical Newton–Einstein universe as itself quantum, as the macroscopic limiting case of the quantum. So, I use ‘quantum’ to include ‘classical’ and by ‘universe’ I mean ‘any universe with a quantum constitution’.

(a)

Omnipresence

‘Omnipresent’ (*def.*) x is omnipresent *iff* x is, and there is not anywhere x is not.

If a minimal realism about the external world is true, quantum events appear not only within the experiments of the physicist. It is empirically well established that physical reality always and everywhere possesses a quantum constitution. Anything quantum is present to the quantum consciousness which pervades it. So, quantum consciousness is not only local but global: there is not anywhere consciousness is not.²⁶

It might seem hard to rule out a priori some existence not dependent on quantum reality and so not dependent on quantum consciousness. But (unlike a physical object or event) consciousness is in itself intrinsically unbounded, so quantum consciousness has that pervasive presence ascribed to God in Psalm 139 8: ‘If I go up to the heavens, you are there; if I make my bed in the depths, you are there’.

(b)

Actus purus

‘Actus purus’ (*def.*) x is actus purus *iff* x is pure actuality.

If anything quantum is present to quantum consciousness, then quantum consciousness is actus purus: pure actuality.²⁷ Distinguish the nunc fluens from the nunc stans, happening, changing, from the time never not being Now.²⁸ Unlike any present event, the Now, or nunc stans, is not bounded or limited. Quantum consciousness as nunc stans actualises quantum events

as the *nunc fluens*. Quantum consciousness is pure actuality (*actus purus*) or the Eternal Now.[29](#)

Quantum consciousness, or anything, entails Presence and, if quantum consciousness is everywhere, there is no *ontological* room for Presence to be the presence of anything other than quantum consciousness.

(c)

Omniscience

‘Omniscient’ (*def.*) *x* is omniscient *iff*, for all *p*, *x* knows that *p*.

If quantum consciousness is omnipresent, it is thereby acquainted with anything quantum anywhere. There is not anything quantum it is not aware of. Intuitively: it *misses* nothing. All the information in the universe or universes is present to it.

It is hard to rule out a priori that there is some non-quantum reality with which quantum consciousness is not acquainted. But, so far as may be discerned by a finite mind, quantum consciousness is omniscient in the propositional sense of knowing the truth values of all the propositions there are, $\forall(p) Kxp$, and so has that property ascribed to God in 1 John 3: 20: ‘God [...] knows all things’.

(d)

Creator

‘Creator’ (*def.*) *x* is the Creator *iff* only *x* makes what is be.

On pain of circularity, whatever caused or causes the space–time universe of physical objects and events, is not spatio-temporal, not physical, and not an event or set of events. Quantum consciousness is at least necessary for the universe and arguably sufficient. It follows that quantum consciousness is not spatio-temporal, not physical, and not an event or set of events. But lacking these features is characteristic of God.

How is this creating done? Quantum consciousness is *actus purus* and pure actuality *actualises*: effects the transition from possible to real, potential to actual. Quantum consciousness created the universe, so has that property ascribed to God in Genesis 1 1: ‘In the beginning, God created the heavens and the earth’.[30](#)

(e)

Sustainer

‘Sustains’ (*def.*) *x* sustains *y* *iff* *y* cannot continue (to be) without *x*.

In a sense, the universe is constantly beginning afresh *now*. Quantum events begin to be *ex nihilo*, and cease to be in *nihilum*. The exploding spontaneity of the present is sustained by quantum consciousness, by, as

we are putting it, non-local consciousness.^{[31](#)} Quantum consciousness actualising explains why anything happens, from moment to moment. Indeed, we would still need quantum consciousness to explain why there is a universe even if there has always been a universe. Consciousness would be required to show why there is an eternal universe rather than no universe whatsoever.

Quantum consciousness sustains the universe in existence from moment to moment, and actualises everything in it by effecting the transition from possibly being to being, from Being to beings.^{[32](#)} Quantum consciousness therefore has the property ascribed to God in Hebrews 1: 3 and Colossians 1: 17: 'In him all things hold together'.

(f)

One

One (*def.*) x is One *iff* x is but does not in principle admit of plurality.

In physics, distinct experiments, on distinct occasions, involve distinct conscious observers. But the universe is a unified multiplicity and any multiplicity may only be unified by unity, by that which is one and only one. Quantum consciousness unifies the universe by being its ground, so quantum consciousness is one and only one. As one, quantum consciousness ontologically resists *duplication*. Quantum consciousness therefore has that property ascribed to God in Deuteronomy 6: 4: 'The Lord is one', Job 23: 13: 'He is unique', Mark 12: 29 'The Lord our God is one Lord', and Mark 12: 32: 'He is one'.

(g)

Simplicity

'Simple' (*def.*) x is simple *iff* x has no parts.

Consciousness has no parts, so quantum consciousness is simple and therefore has that property ascribed to God above in being one. As one, quantum consciousness *ontologically* resists *division*.

(h)

Infinity

'Infinite' (*def.*) x is infinite *iff*, if x is F , there is no limit to the degree to which x is F .

Although the Einstein universe is spatio-temporally finite but unbounded, the universe is in other ways infinite. The universe is mathematical (and so may be described by physics). Mathematics contains infinities. Therefore, so also, in a sense, does the universe. The universe contains an infinite number of facts or truths. The ground of anything infinite is itself infinite. Quantum consciousness is the ground of the universe. Therefore, quantum

consciousness is infinite.[33,34](#) Quantum consciousness therefore possesses the property ascribed to God in Psalm 147: 5 of being 'beyond measure' or, in Ephesians 1: 19, that of having 'infinite might'.[35](#)

(i)

Timelessness

'Timeless' (*def.*) x is timeless *iff* no temporal predicates apply to x .

Quantum reality grounds space–time and quantum consciousness grounds quantum reality. Therefore, quantum consciousness grounds space–time. On pain of circularity, temporal predicates do not apply to the cause of time. Quantum consciousness is eternal, not in lasting infinite time, not in being 'sempiternal', but in *not being in time at all*. Quantum consciousness therefore possesses that property ascribed to God in Deuteronomy 33: 37, 1 Timothy 1: 17, of being 'eternal' or, in Romans 1: 20, of having 'eternal power and character'.

(j)

Good

'Good' (*def.*) x is good *iff* x is conducive to happiness or (*vel*) the elimination of suffering.

Prima facie, quantum consciousness could be morally neutral or evil. However, that there is a universe is better than if there were no universe, and the cause of a good universe is itself good. Quantum consciousness is the cause of the universe, so quantum consciousness is itself good.[36](#) Admittedly, the existence of the universe is a necessary condition for evil and ugliness, but it is also necessary for good and beautiful things. It is better that both good and evil should exist than that no good exists. Quantum consciousness therefore has the property ascribed to God in Psalm 34: 8, Mark 10: 18, of being 'good' and is at least consistent with this rather utilitarian motive in Matthew 5: 45: 'so that you may be sons of your Father who is in heaven; [...] He causes His sun to rise on the evil and the good, and sends rain on the righteous and the unrighteous'.

(k)

Necessary Being

The existence of any particular being looks contingent. It is hard to justify prefacing the existential quantifier with ' $\square$ ' or to find a value of ' x ' that would allow construing any inference from

$$\exists(x) (Fx)$$

to

$$\square\exists(x) (Fx)$$

as valid.

But, if God is Being, rather than 'a' being, then God is necessary because Being cannot not be. 'There is nothing' (in the sense of 'Nothing has existence') is incoherent. Necessarily, there is either Being or Nothing. So, there is necessarily Being. If God is Fundamental, then God is Being or, more accurately, Being is the Being of God.[37](#)

Quantum consciousness, or anything, entails Being and, if quantum consciousness is everywhere, there is no *ontological* room for Being to be the being of anything other than quantum consciousness. Primordial Being is *l'être pour soi*. Quantum consciousness therefore has that property ascribed to God by God in Exodus 3: 14: 'I am that I am'.[38](#)

This 'am' entails but exceeds any 'is'.

Being cannot not be, and so Being is uncreated, and so has the property ascribed to God in Psalm 90: 2: 'Before the mountains were born Or You gave birth to the earth and the world, Even from everlasting to everlasting, You are God'.

If these arguments are sound, quantum consciousness is the unique, infinitely powerful, infinitely knowing, infinitely benevolent, necessary, omnipresent, atemporal ground of everything, a person, and Creator and sustainer of the universe: pure actuality. These are all and only the essential properties ascribed to God within the world's great monotheistic religions. It follows that quantum consciousness is God. Fundamentally understood, God is Being, Presence, Consciousness: the Holy Trinity.[39](#)

God is not an extravagant postulate but an ineliminable presupposition.[40](#) If the quantum universe exists, God exists. A Theory of Everything, worthy of the name, entails a theology.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments

I am grateful to the following for useful discussion over the years: A. J. Ayer, Fergus Kerr, Benedikt Paul Goecke, Saul Kripke, John Leslie, Adrian Moore, Thomas Nagel, Robert Nozick, Terence Penelhum, Roger Penrose, Karl Popper, Mikolaj Slawkowski-Rode, Richard Swinburne, Nicholas Waghorn, Ralph Stefan Weir, Timothy Williamson, John Wisdom.

Conflicts of Interest

The author declares no conflict of interest.

Nomenclature

p, q, r	propositional variables
$\&$	conjunction
$-$	negation
$\rightarrow$	material implication
$\equiv$	biconditional
x, y	name variables
$\exists (x)$	existential quantifier
$\forall (x)$	universal quantifier
$\exists!$	There exists at least and at most one ...
a, b, c	name letters
F, G	predicate letters
$\diamond$	‘It is possible that’
$\square$	‘It is necessary that’

w	possible world
$\in$	‘is a member of’
x	object variable
Φ	particle
$\mathbf{p}$	momentum
$\mathbf{q}$	position
t	time
ψ	wave
Ψ	wave function
F	force
C	‘is a causally sufficient condition for’
IC	‘is an initial condition for’
n	‘number’
NC	‘is a causally necessary condition for’
K	‘It is known that’
B	‘It is believed that’
Θ	‘observes’
$\odot$	‘is conscious’
i	information
I	‘is intelligent’
O	‘occurs’
$P/$	‘the probability of’
P	projection operator on a Hilbert space
H	history
$\mathcal{H}$	Hilbert space
$\mathbf{v}$	vector

Notes

- 1 [Einstein et al. \(1935\)](#). Einstein said: “*Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns nicht näher. Jedenfalls bin ich überzeugt, daß der nicht würfelt.* (The theory delivers much, but it brings us no nearer to the mystery of the Old One. I, at any rate, am convinced that He does not play dice.)” ([Einstein \[1926\] 1971, p. 91](#)). (Translation amended.) The theory is quantum mechanics. The Old One is God. Although the experimenters set out with the explicit aim of finding the classical determinants of quantum reality, Einstein had, in the end, to renounce this scepticism about quantum indeterminism. John Bell showed mathematically that there are no local hidden variables which would eliminate the intrinsically indeterminate and probabilistic nature of quantum entanglement, and its threat of action at a distance and even transmission of information faster than the speed of light. Bell’s refutation of Einstein is confirmed by the experimental work of the 2022 Nobel Prize winners, Alain Aspect, John F. Clauser, and Anton Zeilinger (See [Bell 1964](#)).
- 2 Allocate the value ‘consciousness’ to ‘a’, ‘quantum level’ to ‘b’, and ‘physical world’ to ‘c’, and read ‘if not *x* then not *y*’ as ‘*x* is a necessary condition for *y*’. If not *a* then not *b*, and if not *b* then not *c*, but not: if not *c* then not *b*, then not: if not *b* then not *a*. Otherwise: if not *c* then not *b* (because if not *c* then not *a*, and if not *a* then not *b*, then if not *c* then not *b*, by the transitivity of ‘if not’). But ‘If not *c* then not *b*’ is ruled out by Bell’s criticism of EPR.
- 3 It is sometimes put to me, by way of objection: “Some of the presuppositions you identify are not peculiar to quantum physics but are presuppositions of any theory, or even presuppositions of *anything*”. The observation is correct, but does not count as an objection to ‘Consciousness is a presupposition of quantum physics’. If *p* is a presupposition of any theory, or of anything, then *p* is a presupposition of quantum physics.
- 4 I therefore cannot endorse the central thesis of [Dennett \(2017\)](#), because, if consciousness is a product of evolution, then consciousness is an emergent property of the brain. But consciousness is not an emergent property of the brain, so consciousness is not a product of evolution. Nevertheless, if the thesis that quantum reality presupposes consciousness is correct, then consciousness is ‘already there’ if atoms exist and, given that extra premise, it is right after all that the existence of atoms is sufficient for the existence of consciousness. But then it would remain wrong to claim that consciousness is any kind of product of evolution. Consciousness exists *anyway*.

- 5 Bohr says ‘Far from containing any arbitrary renunciation of customary physical explanation, the notion of complementarity refers directly to our position as observers in a domain of experience where unambiguous application of the concepts used in the description of phenomena depends essentially on the conditions of observation’ ([Bohr 1958, p. 99](#)).
- 6 This claim requires more clarification and defence than it has received. For example: In MW, the logical possibility that MW is false is not realised, neither is the theoretical possibility that the Copenhagen interpretation is correct or the mathematical possibility that there exists only one universe. Moreover: suppose *p* is possible, and *not p* is possible, but *both p and not p* is impossible. Is *p* then realised? Or is *not p* then realised? In MW, there seems nothing to force this one way or the other. Perhaps God does play dice.
- 7 Here we may adduce all Sartre’s grounds for the past being ‘personal’, i.e., indexed to some present consciousness. ([Priest 2001, 163 ff.](#))
- 8 Olley Pearson has argued, in rigorous detail, that a tensed account of time is (a) rationally unavoidable and (b) entails that the self is non-physical. See ([Pearson 2018](#)).
- 9 Einstein says: ‘The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not to individual systems’ Einstein, in ([Schlipp 1970, pp. 671–72](#)).
- 10 For persuasive arguments for intuitionism, see: ([Dummett 1975](#); [Wright 1982, 2004](#)).
- 11 Maudlin points out that: ‘Bohm’s theory is deeply congenial to an ontology which maintains that all that exists exists now, i.e., at a point of time classically conceived’ ([Maudlin 2002, p. 124](#)).
- 12 Paul Davies claims: ‘It is not a wave of any substance or physical stuff, but a wave of knowledge or information. It is a wave that tells us what can be known about the atom, not a wave of the atom itself. Nobody is suggesting that an atom can ever spread itself around like an undulation. But what can spread itself around is what an observer can know about the atom’ ([Davies 1983, p. 107](#)).
- 13 If this is right, does matter exist? Feynman says: ‘You can’t say A is made of B or vice versa. All mass is interaction’. Richard Feynman Note (c. 1948), quoted in ([Gleick 1992, pp. 5, 283](#)).

- 14 Rovelli says ‘If I observe an electron at a certain position, I cannot conclude that the electron is there: I can only conclude that the electron as seen by me is there’ ([Rovelli 2005, p. 116](#)).
- 15 Rovelli thinks ‘the “observer system” *O* in quantum mechanics need not be human or have any other property beside the possibility of interacting with the “observed” system *S*’ ([Rovelli 2005, p. 120](#)). But, as we have seen, any chain of non-conscious observers has to be arrested in a conscious observer or there are no observers.
- 16 One might object that a camera (for example) has a perspective but is not conscious. But it is only right that a camera (or similar) has a perspective if someone could, in principle look through it (or, perhaps, through some camera or other). Consciousness is, therefore, presupposed. This holds even if a given camera can be triggered automatically or with no-one looking through it.
- 17 Rovelli gives ‘The Reason for the Apparent Paradoxes’. He says ‘The incorrect notion that generates the uneasiness with quantum mechanics is the notion of observer-independent state of a system, or observer independent values of physical quantities’ ([Rovelli 2005, p. 120](#)).
- 18 See Note 14 above.
- 19 Rovelli draws a distinction between ‘the measured system *S* and the measuring system *O*’ (116) and says ‘[...] the central tenet of relational quantum mechanics is that there is no meaning in saying that a certain variable of the system *S* takes the value *q*: rather, there is meaning in saying that this variable takes the value *q* for *O*, or with respect to *O*’ ([Rovelli 2005, p. 116](#)).
- 20 For example, Rovelli discusses ‘this relativisation of actuality’ ([Rovelli 2005, p. 116](#)).
- 21 Rovelli endorses ‘Shannon’s definition of information’ and says ‘According to this definition, the statement that *O* has information about *S* actually means that we shall observe *O* and *S* in a subset of the set formed by the Cartesian product [‘OF’chECK] the possible states of *O* and the possible states of *S*’ ([Rovelli 2005, p. 120](#)).
- 22 Rovelli says ‘I reformulate the problem of the interpretation of quantum mechanics as the problem of deriving the formalism from a set of simple physical postulates. I consider a reformulation of quantum mechanics in terms of information theory. All systems are assumed to be equivalent, there is no observer–observed distinction, and the theory describes only the information that systems have about each other; nevertheless, the theory is complete’ ([Rovelli 2005, p. 119](#)).

23	The Bayesian interpretation was developed by Christopher Fuchs, Carlton Caves, and Rüdiger Schack.
24	I argued that classical science cannot explain the scientist presupposed by its practice, in a talk given in Trinity College, Cambridge, in 1988. I thank Mike Brearley, Adrian Moore, and John Wisdom for comments. It was written up as Priest, “Scientist”. (Priest 1993).
25	All and only those logical inferences with Fundamental Ontological import are <i>ontological</i> .
26	Penrose says ‘It seems to me that consciousness is something global’ (Penrose 1997, p. 133). Penrose says ‘I share the discomfort of many others that the “Wigner view” seems to imply a markedly different physical behaviour in our own small corner of the universe from that which would be taking place almost everywhere else, in the absence of local conscious observers to keep things under control’ (Penrose 1987, p. 110).
27	‘Pure act’ is not a perspicacious translation, because it implies that God exists in time. But the immanence of God is the Eternal Now, the unchanging present, the nunc stans. The nunc fluens is Creation.
28	Shimony says: ‘I share Roger [Penrose]’s scepticism about all attempts to interpret away this mystery, for example, by many worlds interpretations, decoherence, hidden variables, etc. At some stage or another in a measurement process the unitary evolution of the quantum state breaks down and an actualization occurs. But at what stage? There are many possibilities’ (Shimony 1997, pp. 156–57).
29	Schrödinger says ‘Mind [...] I venture to call it indestructible since it has a peculiar timetable, namely mind is always <i>now</i> . There is really no before or after for mind. There is only a now that includes memories and expectations. But I grant that our language is not quite adequate to express this, and I also grant, should anyone wish to state it, that I am now talking religion, not science’ Schrödinger (QQ: 87)
30	Interestingly, it follows that consciousness is not a product of evolution. Consciousness does not evolve.
31	Davies says ‘Science, which by definition deals only with the physical universe, might successfully explain one thing in terms of another, and that in terms of another and so on, but the totality of physical things requires an explanation from without’ (Davies 1983, p. 47).
32	Omnès says ‘In his [Heidegger’s] opinion, one cannot reach [the] Being by experiment, because [the] “Being withdraws itself as soon as it is disclosed in the

	instant”. Whatever the meaning of this poetic sentence, it strangely looks when reading it in its context like a transmutation from potentiality to actuality that is not so foreign to physics’ (Omnès 1994, pp. 519–20).
33	Rees says ‘It is indeed remarkable that the external world displays so many patterns that our minds can interpret in mathematical “language”—especially when these are so remote from the everyday experiences and phenomena that our brains have evolved to cope with’ (Rees 1997, p. 160).
34	Planck says ‘All matter originates and exists only by virtue of a force which brings the particle of an atom to vibration and holds this most minute solar system of the atom together. We must assume behind this force the existence of a conscious and intelligent mind. This mind is the matrix of all matter’ (Planck 1944 , Abt. Va, Rep. 11, Nr. 1797).
35	Heisenberg says ‘If we wish to approach the “one” in terms of a precise scientific language, we must turn our attention to that center of science described by Plato, in which the fundamental mathematical symmetries are to be found. In the concepts of this language we must be content with the statement that “God is a mathematician”; for we have freely chosen to confine our vision to that realm of being which can be understood in the mathematical sense of the word “understanding”, which can be described in rational terms’ (Werner Heisenberg 1985, p. 54).
36	It has been argued persuasively by the Canadian philosopher John Leslie, in a rather Neo-Platonic way, that <i>the reason why</i> there is a universe is that it is better that there should be a universe rather than no universe. See (Leslie 1979).
37	Here ‘Being’ does not mean ‘the totality of what is’ (or similar). ‘Being’ is understood as ‘Pure Being’. Being is the Being of Parmenides and sharply <i>ontologically</i> distinct from any or all beings.
38	It follows that the Thomist thesis that God is his own existence is correct. (Aquinas 1953, VII. II ; 1993, 202 & ff.) “Oportet ergo quod hoc quod est esse, sit substantia vel natura Dei”, “Therefore that which is Being, must be the essence or nature of God” (translation amended.) VII. II. 11.
39	Roger Penrose comes close to expressing this doctrine and (no doubt unknowingly) engages in quasi-Hegelian dialectic, in (Penrose 1997, pp. 137–39). He uses the image of a triangle (138) to describe the relations between three ‘worlds’ (96) and three ‘mysteries’. [1] Platonic; [2] Mental; [3] Physical. Penrose argues that each ‘emerges’ from the other. Penrose says ‘There is something distinctly mysterious about the way that these three worlds interrelate with one another—where each

seems almost to “emerge” from a small part of its predecessor’ ([Penrose 1997, p. 139](#)).

- 40 Einstein notoriously resisted the more counter-intuitive implications of quantum physics for years. Nevertheless, Einstein understands, better than most, that serious science presupposes consciousness. Einstein says ‘[...] everyone who is seriously involved in the pursuit of science becomes convinced that a spirit is manifest in the laws of the universe—a spirit vastly superior to that of man, and one in the face of which we with our modest powers must feel humble’ ([Einstein \[1926\] 1971, p. 33](#)).

References

1. Aquinas, St. Thomas. 1953. *Quaestiones Disputatae de Potentia*. Turin: Marietti. [[Google Scholar](#)]
2. Aquinas, St. Thomas. 1993. *Selected Philosophical Writings*. Translated by Timothy McDermott. Oxford: Oxford University Press. [[Google Scholar](#)]
3. Bell, John S. 1964. On the Einstein, Podolsky, Rosen Paradox. *Physics* 1: 195–200. [[Google Scholar](#)] [[CrossRef](#)]
4. Bohr, Niels. 1958. *Atomic Physics and Human Knowledge*. New York: John Wiley. [[Google Scholar](#)]
5. Davies, Paul. 1983. *God and the New Physics*. London: Dent. [[Google Scholar](#)]
6. Dennett, Daniel C. 2017. *From Bacteria to Bach and Back: The Evolution of Minds*. New York: Norton. [[Google Scholar](#)]
7. De Witt, Bryce S. 1970. Quantum Mechanics and Reality. *Physics Today* 23: 30–35. [[Google Scholar](#)] [[CrossRef](#)]
8. Dummett, Michael. 1975. The Philosophical Basis of Intuitionistic Logic. In *Logic Colloquium 73*. Special Issue of *Studies in Logic and the Foundations of Mathematics*. Edited by H. E. Rose and J. C. Shepherdson. New York: North-Holland, vol. 80, pp. 5–40. [[Google Scholar](#)]
9. Einstein, A., B. Podolsky, and N. Rosen. 1935. Can Quantum Mechanical Description of Physical reality be Considered Complete? *Physical Review* 47: 777–80. [[Google Scholar](#)] [[CrossRef](#)]
10. Einstein, Albert. 1971. Letter to Max Born. In *The Born-Einstein Letters*. Translated by Irene Born. New York: Walker and Coy. First published 1926. [[Google Scholar](#)]
11. Gleick, James. 1992. *Genius: The Life and Science of Richard Feynman*. New York: Open Road Media. [[Google Scholar](#)]

12. Griffiths, Robert. 1984. *Consistent histories and the Interpretation of Quantum Mechanics*. Berlin: Springer. [[Google Scholar](#)]
13. Leslie, John. 1979. *Value and Existence*. Oxford: Blackwell. [[Google Scholar](#)]
14. Maudlin, Tim. 2002. *Quantum Non-Locality and Relativity*, 2nd ed. Oxford: Blackwell. [[Google Scholar](#)]
15. Omnès, Roland. 1994. *The Interpretation of Quantum Mechanics*. Princeton: Princeton University Press. [[Google Scholar](#)]
16. Pearson, Olley. 2018. *Rationality, Time, and Self*. London: Palgrave. [[Google Scholar](#)]
17. Penrose, Roger. 1987. Quantum Physics and Conscious Thought. In *Quantum Implications: Essays in Honour of David Bohm*. Edited by Basil J. Hiley and David Peat. London: Routledge. [[Google Scholar](#)]
18. Penrose, Roger. 1997. *The Large, the Small and the Human Mind*. Oxford: Oxford University Press. [[Google Scholar](#)]
19. Planck, Max. 1944. 'Das Wesen der Materie' ['The Nature of Matter'] *Lecture Delivered in Florence, Italy*. Berlin: Archiv zur Geschichte der Max-Planck-Gesellschaft, Abt. Va, Rep. 11 Planck, Nr. 1797. [[Google Scholar](#)]
20. Priest, Stephen. 1993. Newton and Hegel: Can Science Explain the Scientist? In *Hegel and Newtonianism*. Edited by Michael J. Petry. The Hague: Kluwer. [[Google Scholar](#)]
21. Priest, Stephen, ed. 2001. *Jean-Paul Sartre: Basic Writings*. London: Routledge. [[Google Scholar](#)]
22. Rees, Martin. 1997. *Before the Beginning: Our Universe and Others*. Reading Mass: Addison-Wesley. [[Google Scholar](#)]
23. Rovelli, Carlo. 2005. Relational Quantum Mechanics. In *Quo Vadis Quantum Mechanics?* Edited by Avshalom Elitzur, Shoham Dolev and Nancy Kolenda. With a Forward by Roger Penrose. Berlin: Springer. [[Google Scholar](#)]
24. Schlipp, Paul A. 1970. *Albert Einstein: Philosopher-Scientist*. New York: Harper & Row. [[Google Scholar](#)]
25. Shimony, Abner. 1997. *On Mentality, Quantum Mechanics and the Actualisation of Potentialities*. Penrose. Cambridge: Cambridge University Press. [[Google Scholar](#)]
26. Werner Heisenberg. 1985. Consequences for the Evolution of Human Thought in Our Own Day. In *Quantum Questions*. Edited by Ken Wilber. Shambala: Boulder Co., pp. 52–54. [[Google Scholar](#)]

27. Wright, Crispin. 1982. Strict Finitism. *Synthese* 51: 203–82. [[Google Scholar](#)] [[CrossRef](#)]
28. Wright, Crispin. 2004. Intuition, Entitlement and the Epistemology of Logical Laws. *Dialectica* 58: 155–75. [[Google Scholar](#)] [[CrossRef](#)]

A New Theory in Physics Claims to Solve the Mystery of Consciousness

·August 11, 2022

Summary: *Consciousness can not simply be reduced to neural activity alone, researchers say. A novel study reports the dynamics of consciousness may be understood by a newly developed conceptual and mathematical framework.*

Source: Bar-Ilan University

How do 1.4 kg of brain tissue create thoughts, feelings, mental images, and an inner world?

The ability of the brain to create consciousness has baffled some for millennia. The mystery of consciousness lies in the fact that each of us has subjectivity, something that is like to sense, feel and think.

In contrast to being under anesthesia or in a dreamless deep sleep, while we're awake we don't "live in the dark" — we experience the world and ourselves. But how the brain creates the conscious experience and what area of the brain is responsible for this remains a mystery.

According to Dr. Nir Lahav, a physicist from Bar-Ilan University in Israel, "This is quite a mystery since it seems that our conscious experience cannot arise from the brain, and in fact, cannot arise from any physical process."

As strange as it sounds, the conscious experience in our brain, cannot be found or reduced to some neural activity.

"Think about it this way," says Dr. Zakaria Neemeh, a philosopher from the University of Memphis, "when I feel happiness, my brain will create a distinctive pattern of complex neural activity. This neural pattern will perfectly correlate with my conscious feeling of happiness, but it is not my actual feeling. It is just a neural pattern that represents my happiness. That's why a scientist looking at my brain and seeing this pattern should ask me what I feel, because the pattern is not the feeling itself, just a representation of it."

As a result, we can't reduce the conscious experience of what we sense, feel and think to any brain activity. We can just find correlations to these experiences.

After more than 100 years of neuroscience we have very good evidence that the brain is responsible for the creation of our conscious abilities. So how could it be that these conscious experiences can't be found anywhere in the brain (or in the body) and can't be reduced to any neural complex activity?

This mystery is known as the hard problem of consciousness. It is such a difficult problem that until a couple of decades ago only philosophers discussed it and even today, although we have made huge progress in our understanding of the neuroscientific basis of consciousness, still there is no adequate theory that explains what consciousness is and how to solve this hard problem.

Dr. Lahav and Dr. Neemeh recently published a new physical theory in the journal *Frontiers in Psychology* that claims to solve the hard problem of consciousness in a purely physical way.

According to the authors, when we change our assumption about consciousness and assume that it is a relativistic phenomenon, the mystery of consciousness naturally dissolves. In the paper the researchers developed a conceptual and mathematical framework to understand consciousness from a relativistic point of view.

According to Dr. Lahav, the lead author of the paper, “consciousness should be investigated with the same mathematical tools that physicists use for other known relativistic phenomena.”

To understand how relativity dissolves the hard problem, think about a different relativistic phenomenon, constant velocity. Let's choose two observers, Alice and Bob, where Bob is on a train that moves with constant velocity and Alice watches him from the platform. there is no absolute physical answer to the question what the velocity of Bob is.

The answer is dependent on the frame of reference of the observer.

From Bob's frame of reference, he will measure that he is stationary and Alice, with the rest of the world, is moving backwards. But from Alice's frame Bob is the one that's moving and she is stationary.

Although they have opposite measurements, both of them are correct, just from different frames of reference.

Because, according to the theory, consciousness is a relativistic phenomenon, we find the same situation in the case of consciousness.

Now Alice and Bob are in different cognitive frames of reference. Bob will measure that he has conscious experience, but Alice just has brain activity with no sign of the actual conscious experience, while Alice will measure that she is the one that has consciousness and Bob has just neural activity with no clue of its conscious experience.

Just like in the case of velocity, although they have opposite measurements, both of them are correct, but from different cognitive frames of reference.

As a result, because of the relativistic point of view, there is no problem with the fact that we measure different properties from different frames of reference.

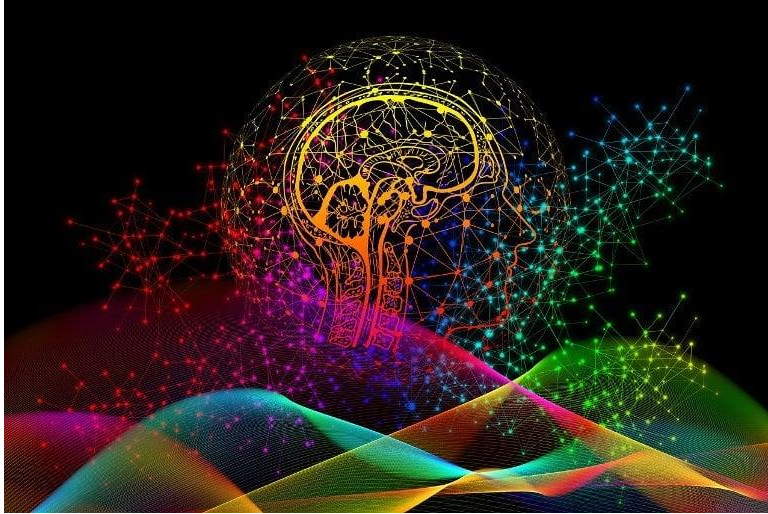
The fact that we cannot find the actual conscious experience while measuring brain activity is because we're measuring from the wrong cognitive frame of reference.

According to the new theory, the brain doesn't create our conscious experience, at least not through computations. The reason that we have conscious experience is because of the process of physical measurement.

In a nutshell, different physical measurements in different frames of reference manifest different physical properties in these frames of reference although these frames measure the same phenomenon.

For example, suppose that Bob measures Alice's brain in the lab while she's feeling happiness. Although they observe different properties, they actually measure the same phenomenon from different points of view. Because of their different kinds of measurements, different kinds of properties have been manifested in their cognitive frames of reference.

For Bob to observe brain activity in the lab, he needs to use measurements of his sensory organs like his eyes. This kind of sensory measurement manifests the substrate that causes brain activity – the neurons.



After more than 100 years of neuroscience we have very good evidence that the brain is responsible for the creation of our conscious abilities. Image is in the public domain

Consequently, in his cognitive frame Alice has only neural activity that represents her consciousness, but no sign of her actual conscious experience itself. But, for Alice to measure her own neural activity as happiness, she uses different kind of measurements. She doesn't use sensory organs, she measures her neural representations directly by interaction between one part of her brain with other parts. She measures her neural representations according to their relations to other neural representations.

This is a completely different measurement than what our sensory system does and, as a result, this kind of direct measurement manifests a different kind of physical property. We call this property conscious experience.

As a result, from her cognitive frame of reference, Alice measures her neural activity as conscious experience.

Using the mathematical tools that describe relativistic phenomena in physics, the theory shows that if the dynamics of Bob's neural activity could be changed to be like the dynamics of Alice's neural activity, then both will be in the same cognitive frame of reference and would have the exact same conscious experience as the other.

Now the authors want to continue to examine the exact minimal measurements that any cognitive system needs in order to create consciousness.

The implications of such a theory are huge. It can be applied to determine which animal was the first animal in the evolutionary process to have consciousness, when a fetus or baby begins to be conscious, which patients with consciousness disorders are conscious, and which AI systems already today have a low degree (if any) of consciousness.

About this consciousness and physics research news

Author: [Elana Oberlander](#)

Source: [Bar-Ilan University](#)

Contact: Elana Oberlander – Bar-Ilan University

“[A Relativistic Theory of Consciousness](#)” by Nir Lahav et al. *Frontiers in Psychology*

Abstract

A Relativistic Theory of Consciousness

In recent decades, the scientific study of consciousness has significantly increased our understanding of this elusive phenomenon. Yet, despite critical development in our understanding of the

functional side of consciousness, we still lack a fundamental theory regarding its phenomenal aspect.

There is an “explanatory gap” between our scientific knowledge of functional consciousness and its “subjective,” phenomenal aspects, referred to as the “hard problem” of consciousness. The phenomenal aspect of consciousness is the first-person answer to “what it’s like” question, and it has thus far proved recalcitrant to direct scientific investigation.

Naturalistic dualists argue that it is composed of a primitive, private, non-reductive element of reality that is independent from the functional and physical aspects of consciousness. Illusionists, on the other hand, argue that it is merely a cognitive illusion, and that all that exists are ultimately physical, non-phenomenal properties.

We contend that both the dualist and illusionist positions are flawed because they tacitly assume consciousness to be an absolute property that doesn’t depend on the observer.

We develop a conceptual and a mathematical argument for a relativistic theory of consciousness in which a system either has or doesn’t have phenomenal consciousness *with respect to some observer*.

Phenomenal consciousness is neither private nor delusional, just relativistic. In the frame of reference of the cognitive system, it will be observable (first-person perspective) and in other frame of reference it will not (third-person perspective). These two cognitive frames of reference are both correct, just as in the case of an observer that claims to be at rest while another will claim that the observer has constant velocity.

Given that consciousness is a relativistic phenomenon, neither observer position can be privileged, as they both describe the same underlying reality. Based on relativistic phenomena in physics we developed a mathematical formalization for consciousness which bridges the explanatory gap and dissolves the hard problem.

Given that the first-person cognitive frame of reference also offers legitimate observations on consciousness, we conclude by arguing that philosophers can usefully contribute to the science of consciousness by collaborating with neuroscientists to explore the neural basis of phenomenal structures.

A Relativistic Theory of Consciousness

[Nir Lahav](#)^{1*} [Zachariah A. Neemeh](#)^{2,3}

- ¹Department of Physics, Bar-Ilan University, Ramat Gan, Israel
- ²Department of Philosophy, The University of Memphis, Memphis, TN, United States
- ³Institute for Intelligent Systems, The University of Memphis, Memphis, TN, United States

In recent decades, the scientific study of consciousness has significantly increased our understanding of this elusive phenomenon. Yet, despite critical development in our understanding of the functional side of consciousness, we still lack a fundamental theory regarding its phenomenal aspect. There is an “explanatory gap” between our scientific knowledge of functional consciousness and its “subjective,” phenomenal aspects, referred to as the “hard problem” of consciousness. The phenomenal aspect of consciousness is the first-person answer to “what it’s like” question, and it has thus far proved recalcitrant to direct scientific investigation. Naturalistic dualists argue that it is composed of a primitive, private, non-reductive element of reality that is independent from the functional and physical aspects of consciousness. Illusionists, on the other hand, argue that it is merely a cognitive illusion, and that all that exists are ultimately physical, non-phenomenal properties. We contend that both the dualist and illusionist positions are flawed because they tacitly assume consciousness to be an absolute property that doesn’t depend on the observer. We develop a conceptual and a

mathematical argument for a relativistic theory of consciousness in which a system either has or doesn't have phenomenal consciousness *with respect to some observer*. Phenomenal consciousness is neither private nor delusional, just relativistic. In the frame of reference of the cognitive system, it will be observable (first-person perspective) and in other frame of reference it will not (third-person perspective). These two cognitive frames of reference are both correct, just as in the case of an observer that claims to be at rest while another will claim that the observer has constant velocity. Given that consciousness is a relativistic phenomenon, neither observer position can be privileged, as they both describe the same underlying reality. Based on relativistic phenomena in physics we developed a mathematical formalization for consciousness which bridges the explanatory gap and dissolves the hard problem. Given that the first-person cognitive frame of reference also offers legitimate observations on consciousness, we conclude by arguing that philosophers can usefully contribute to the science of consciousness by collaborating with neuroscientists to explore the neural basis of phenomenal structures.

Introduction

The Hard Problem of Consciousness

As one of the most complex structures we know of nature, the brain poses a great challenge to us in understanding how higher functions like perception, cognition, and the self arise from it. One of its most baffling abilities is its capacity for conscious experience ([van Gulick, 2014](#)). Thomas [Nagel \(1974\)](#) suggests a now widely accepted definition of consciousness: a being is conscious just if there is “something that it is like” to be that creature, i.e., some subjective way the world seems or appears from the creature's point of view. For example, if bats are conscious, that means there is something it is like for a bat to experience its world through its echolocational senses. On the other hand, under deep sleep (with no dreams) humans are unconscious because there is nothing it is like for humans to experience their world in that state.

In the last several decades, consciousness has transformed from an elusive metaphysical problem into an empirical research topic. Nevertheless, it remains a puzzling and thorny issue for science. At the heart of the problem lies the question of the brute phenomena that we experience from a first-person perspective—e.g., what it is like to feel redness, happiness, or a thought. These qualitative states, or qualia, compose much of the phenomenal side of consciousness. These qualia are arranged into spatial and temporal patterns and formal structures in phenomenal experience, called eidetic or transcendental structures¹. For example, while qualia pick out how a specific note sounds, eidetic structures refer to the temporal form of the whole melody. Hence, our inventory of the elusive properties of phenomenal consciousness includes both qualia and eidetic structures.

One of the central aspects of phenomenal features is privacy. It seems that my first-person feeling of happiness, for example, cannot be measured from any other third-person perspective. One can take indirect measurements of my heart rate or even measure the activity in the brain networks that create the representation of happiness in

my mind, but these are just markers of the feeling that I have, and not the feeling itself ([Block, 1995](#)). From my first-person perspective, I don't feel the representation of happiness. I just feel happiness. Philosophers refer to this feature of phenomenal consciousness as "transparency": we seem to directly perceive things, rather than mental representations, even though mental representations mediate experience. But this feeling cannot be directly measured from the third-person perspective, and so is excluded from scientific inquiry. Yet, let us say that we identify a certain mental representation in a subject's brain as "happiness." What justifies us in calling it "happiness" as opposed to "sadness?" Perhaps it is correlated with certain physiological and behavioral measures. But, ultimately, the buck stops somewhere: at a subject's phenomenological report otherwise, we cannot know what this representation represents ([Gallagher and Zahavi, 2020](#)). If we categorize the representation as "happiness" but the subject insists they are sad, we are probably mistaken, not the subject. Phenomenal properties are seemingly private and, by some accounts, even beyond any physical explanation. For being physical means being public and measurable and is epistemically inconsistent with privacy. Take an electron, for example. We know it is physical because we can measure it with e.g., a cathode ray tube. If electrons were something that only I could perceive (i.e., if they were private), then we would not include them as physical parts of the scientific worldview. Affective scientists, for example, measure aspects of feeling all the time, like valence and arousal, but measuring these is not the same as measuring the feeling of happiness itself. What affective scientists measure are outcomes of the feeling, while the feeling itself is private and not measurable by the scientists. To date, it is not clear how to bridge the "explanatory gap" ([Levine, 1983](#)) between "private" phenomenal features and public, measurable features (e.g., neurocomputational structures), leaving us stuck in the "hard problem" of consciousness ([Chalmers, 1996](#)). This gap casts a shadow on the possibility for neuroscience to solve the hard problem because all explanations will always remain within a third-person perspective (e.g., neuronal firing patterns and representations), leaving the first-person perspective out of the reach of neuroscience. This situation divides consciousness into two separate aspects, the functional aspect and the phenomenal one ([Block, 1995](#)). The functional aspect ('functional consciousness'), is the objectively observable aspect of consciousness ([Franklin et al., 2016](#); [Kanai et al., 2019](#)). [In that sense, it's similar to Ned Block's definition of access conscious, but with less constraints. All phenomenal consciousness has a functional aspect and vice versa, whereas for Block this strict equivalence doesn't hold ([Block, 2011](#)).] But the subjective aspect (phenomenal consciousness) is not directly observable except on the part of the person experiencing that conscious state. As we saw above, because they are private, phenomenal properties are distinct from any cognitive and functional property (which can be publicly observable). Any theory of consciousness should explain how to bridge this gap—How can functional, public, aspects give rise to phenomenal, private aspects?

Nevertheless, in recent decades, consciousness has become increasingly amenable to empirical investigation by focusing on its functional aspect, finally enabling us to begin to understand this enigmatic phenomenon. For example, we now have good evidence that consciousness doesn't occur in a single brain area. Rather, it seems to be a global

phenomenon in widespread areas of the brain ([Baars, 1988](#); [Varela et al., 2001](#); [Mashour et al., 2020](#)).

In studying *functional consciousness*, we take consciousness to be a form of information processing and manipulation of representations, and we trace its functional or causal role within the cognitive system. Widely successful theories, such as global workspace theory ([Baars, 1988](#); [Dehaene, 2014](#); [Franklin et al., 2016](#)), attention schema theory ([Graziano, 2019](#)), recurrent processing theory ([Lamme, 2006](#); [Fahrenfort et al., 2007](#)), and integrated information theory ([Tononi, 2008](#); [Oizumi et al., 2014](#)) are virtually premised on this information processing account. Despite our advances in the study of functional consciousness, we still lack a convincing way to bridge the explanatory gap to phenomenal consciousness. The questions, “Why does it feel like anything at all to process information?”, and “How can this feeling be private?” still remain controversial. The hard problem primarily remains a philosophical rather than scientific question. Phenomenal consciousness must be the ultimate reference point for any scientific theory of consciousness. Ultimately, theories of consciousness as information processing, i.e., theories of functional consciousness, only approximate full-blown consciousness by abstracting away its phenomenal features. But they must ultimately refer to phenomenal features in order to give a full explanation of consciousness. Otherwise, there are no grounds for labeling them theories of consciousness, rather than theories of cognition or global informational access.

Let’s take Integrated Information Theory (IIT), for example. IIT claims to bridge between the phenomenal and functional aspects of consciousness and to answer the question of what qualia are ([Tononi et al., 2016](#)). According to IIT, consciousness is the result of highly integrated information in the brain [which is “the amount of information generated by a complex of elements, above and beyond the information generated by its parts” ([Tononi, 2008](#))]. A mathematical formula can determine how much integrated information, and thus how much consciousness, is present in any specific system. This measure is referred to as ϕ , and if ϕ is higher than zero, the system is conscious. In this situation, the system has a “maximally irreducible conceptual structure” which is identical to its phenomenal experience ([Oizumi et al., 2014](#)). These maximally irreducible conceptual structures are composed of integrated information states, and thus a quale is identical to a specific relation between integrated information states. The problem is that these maximally irreducible conceptual structures are physical states and even though they can be very complex, they are still public and observable from a third-person perspective, while qualia are private and only perceptible from the first-person perspective. How can there be an identity between opposite properties like public and privacy properties? Yet, this is what IIT suggests without any explanation about how to bridge this contradiction. Let’s choose a conceptual structure such that it is a quale of happiness according to IIT. However, it cannot truly be the quale of happiness because, in principle, when we measure it, we observe a physical process and not the qualitative happiness itself (that’s why it’s not enough to measure the physical process, and eventually we need to ask the participants what they feel). Again, we remain stuck within the explanatory gap and the hard problem. In other words, IIT hasn’t solved the hard problem and hasn’t bridged the explanatory gap between private and public properties ([Mindt, 2017](#)). In order to avoid this problem, IIT needs to assume

phenomenal consciousness as a primitive element separated from other physical elements like space and time. In that case, there is no gap to bridge because now these maximally irreducible conceptual structures are not qualia anymore. They are just physical structures that have properties that correspond to the properties of phenomenal consciousness, but that are not identical to them. In that sense, IIT just shows correspondence between two separate kinds of elements, phenomenal and physical. In the next paragraph we will see that this kind of solution is exactly what Chalmers refers to as ‘naturalistic dualism.’ The problem with that is that instead of understanding what phenomenal consciousness is and what the relations are between the physical world and the phenomenal world, it just assumes that phenomenal consciousness exists as a separate basic element in the world, and it does not solve how there can be any interaction between physical and phenomenal elements ([Carroll, 2021](#)).

This is not an issue of only IIT, but a problem to all physicalist’s theories. As David [Chalmers \(1995\)](#) has put it, “the structure and dynamics of physical processes yield only more structure and dynamics, so structures and functions are all we can expect these processes to explain.” The structure and dynamics of information tell us nothing of the story of how one gets from these public structures and dynamics to our private phenomenal experience (and how there can be an equality between these opposite properties). This is known as the ‘structure and dynamics argument’ ([Alter, 2016](#); [Chalmers, 2003](#); [Mindt, 2017](#)), stating that structure and dynamics alone are not enough to account for consciousness. This raises a concern about whether any physicalist theory can solve the hard problem of consciousness. In the next sections, however, we will show that physicalism is broader than describing only structures and dynamics, and that we can use this fact in order to solve the hard problem.

Views about the relation between phenomenal and functional consciousness exist across a spectrum. On one end, illusionists seek to erase the hard problem by referring only to functional consciousness, taking phenomenal properties to be cognitive illusions. On the other end, naturalistic dualists or panpsychists seek to promote phenomenal consciousness as a fundamental, non-material, and undecomposable constituent of nature ([Chalmers, 2017](#)). The controversy over phenomenal consciousness can be traced to one central problem: naturalization. The project of naturalization involves taking folk psychological concepts and subjecting them to physical laws and empirical scrutiny ([Hutto, 2007](#)). Illusionists take the current scientific approach to consciousness and argue that this eliminates the messy problem with supposedly private, immaterial qualia. According to them, functional consciousness generates an illusion of special phenomenal properties, which create the persistent “user illusion” ([Dennett, 1991](#)) of a first-person perspective. Dualists start from the same problem of naturalization, but take it that phenomenal consciousness is simply not amenable to third-person scientific inquiry due to its sui generis properties. What is needed, according to the naturalistic dualist, is an expanded understanding of what counts as “natural.” “Given that reductive explanation fails, *nonreductive* explanation is the natural choice” ([Chalmers, 2017](#), p. 359). Chalmers proposes that consciousness is a fundamental property, something like the strong nuclear force that is irreducible to other forces. A complete ontology of the

natural world simply must include phenomenal consciousness as a basic, undecomposable constituent.

The Zombie Argument and the Paradox of Phenomenal Judgment

[Chalmers \(1996\)](#) discusses the logical possibility of a *zombie*, a being physically, cognitively, and behaviorally identical to a human, but lacking phenomenal consciousness altogether. One would think that such a creature would be dull, like a robot with basic automated responses, but this is not the case. Because a zombie is physically identical to a human, it means that it has the same cognitive system as us: a system that gets inputs from the environment, processes them, and creates behavior responses. In fact, there are no differences between the human and the zombie's cognitive dynamics, representations, and responses. But, for the zombie, there is nothing that is like to do all these processes. According to [Chalmers \(1996\)](#), the zombie has phenomenal judgments. This concept is very important for our argument, so let's examine it a bit. Phenomenal judgements are higher-order cognitive functions that humans and zombies have in common. Humans are aware of their experience and its contents and they can form judgments about it (e.g., when we think 'There is something red'), then, usually, they are led to make claims about it. These various judgments in the vicinity of consciousness are phenomenal judgments. They are not phenomenal states themselves, but they are about phenomenology. Phenomenal judgments are often reflected in claims and reports about consciousness, but they start as a mental process. Phenomenal judgments are themselves cognitive acts that can be explained by functional aspects like the manipulation of mental representations. That's why zombies also have phenomenal judgements. We can think of a judgment as what is left of a belief after any associated phenomenal property is subtracted. As a result, phenomenal judgements are part of the functional aspect of consciousness. As Chalmers puts it (1996, p.174):

"Judgments can be understood as what I and my zombie twin have in common. A zombie does not have any conscious experience, but he claims that he does. My zombie twin judges that he has conscious experience, and his judgments in this vicinity correspond one-to-one to mine. He will have the same form, and he will function in the same way in directing behavior as mine... Alongside every conscious experience there is a content-bearing cognitive state. It is roughly information that is accessible to the cognitive system, available for verbal report, and so on."

In other words, phenomenal judgments can be described as the representations of a cognitive system that bear content about phenomenology (which are not necessarily linguistic representations. E.g., such representation can be a representation of the color of the apple). In this paper, we will identify functional consciousness with the creation of phenomenal judgments.

As a result of the zombie's capacity to create phenomenal judgements, we reach a peculiar situation: The zombie has functional consciousness, i.e., all the physical and functional conscious processes studied by scientists, such as global informational access. But there would be nothing it is like to have that global informational access and

to be that zombie. All that the zombie cognitive system requires is the capacity to produce phenomenal judgments that it can later report. For example, if you asked it if it sees a red rose in front of it, using information processing, it might respond, “Yes, I’m definitely conscious of seeing a red rose,” although it is ultimately mistaken and there is truly nothing that is like for the zombie to see that rose. In order to produce this phenomenal judgment, despite having no phenomenal consciousness, the zombie cognitive system needs representations and a central system with direct access to important information enabling it to generate behavioral responses. It needs direct access to perceptual information, a concept of self to differentiate itself from the world, an ability to access its own cognitive contents, and the capacity to reflect. Such a cognitive system could presumably reason about its own perceptions. It would report that it sees the red rose, and that it has some property over and above its structural and functional properties—phenomenal consciousness. Of course, this report would be mistaken. It is a paradoxical situation in which functional consciousness creates phenomenal judgments without the intervention of phenomenal consciousness—yet phenomenal judgments are purportedly about phenomenal consciousness. This paradox of phenomenal judgment ([Chalmers, 1996](#)) arises because of the independence of phenomenal consciousness from physical processes. The hidden assumption here is that consciousness is private. Consequently, it is not possible to measure it. It seems that one aspect of consciousness (physical, functional consciousness) can come without the other (phenomenal consciousness). For example, in IIT there is a variant of the phenomenal judgement paradox. According to IIT, there can be a cognitive system that will manipulate information and infer that it has phenomenal experience, that there is something that is like to be that cognitive system, yet it doesn’t have any consciousness because its neural network creates all these judgements in a feed-forward way, meaning that $\phi = 0$ in the system ([Doerig et al., 2019](#)). Again, we see that even in IIT the functional part of consciousness (phenomenal judgements) can come without the phenomenal aspect of consciousness. This paradox arises because in IIT consciousness and cognitive content are not conditioned to correspond to each other. As a result, although the cognitive system has phenomenal judgements about phenomenal consciousness, still, it is just a zombie with no phenomenal consciousness because $\phi = 0$ in the system (this kind of a zombie is termed functional zombie, see [Oizumi et al., 2014](#); [Tononi et al., 2016](#)).

In order to solve this paradox, we need to explain two aspects of consciousness: How there could be natural phenomena that are private and thus independent of physical processes (or how come they *seem* private), and what the exact relationship between cognitive content and phenomenal consciousness is.

The illusionist position is that phenomenal properties are cognitive illusions generated by the brain. If a zombie with developed cognitive abilities can mistakenly think it has phenomenal consciousness, how do we know that this is not the case with ourselves, as well? For the illusionist, this is exactly the predicament we are in, albeit we are zombies with a rich inner life ([Frankish, 2017](#))—whatever that means. The illusionist takes the purported scientific intractability of phenomenal consciousness to be evidence against phenomenal consciousness. “Illusionists deny that experiences have

phenomenal properties and focus on explaining why they seem to have them” ([Frankish, 2017](#), p. 18). While this position might seem to be counterintuitive, it saves a conservative understanding of physics and obviates any call for exotic properties of the universe, as [Chalmers \(2017\)](#) argues for. As [Graziano et al. \(2020\)](#) state:

“‘I know I have an experience because, Dude, I’m experiencing it right now.’ Every argument in favour of the literal reality of subjective experience ... boils down sooner or later to that logic. But the logic is circular. It is literally, ‘X is true because X is true.’ If that is not a machine stuck in a logic loop, we don’t know what is” ([Graziano et al., 2020](#), p. 8).

The problem, however, is that it’s not clear what the claim that phenomenal consciousness is an illusion means. What exactly does “illusion” mean in this context? To be clear, the illusionist is not denying that you see, e.g., a red rose in front of you. You do “see” it, but it only *seems* like it has phenomenal properties. You have non-phenomenal access to the perceptual representation, sufficient to enable a phenomenal judgment about that representation (e.g., “I see a red rose”). This contention runs in the face of all our intuitions, but the illusionist claims those intuitions are illusory. [Frankish \(2017\)](#) states that these *seeming*-properties are quasi-phenomenal properties, which are physical properties that give the illusion of phenomenal properties. But are quasi-phenomenal properties any less mysterious than phenomenal properties? How is it that seeming to be phenomenally conscious is not just being phenomenally conscious?

One way to address these problems and understand the illusionist position is to understand them as agreeing with Chalmers about phenomenal consciousness’s privacy, and that zombies are logically possible. Then, the paradox of phenomenal judgment already notes that just such a functionally conscious cognitive system could produce phenomenal judgments without having phenomenal consciousness. However, illusionists take this to be evidence against phenomenal consciousness. A purely physical system creates phenomenal judgments, therefore there are nothing but purely physical processes involved (and hence, no qualia). However, this position is problematic. Only if phenomenal properties are private could there be such a paradoxical situation of having phenomenal judgments about phenomenal consciousness, yet without any phenomenal consciousness. If we can argue *against* the privacy of phenomenal properties, then we can escape the trap into which both the dualist and illusionist fall.

We interpret the dualist and illusionist extremes as unfortunate consequences of a mistaken view of naturalism. The illusionist’s commitment to naturalism leads them to exclude supposedly non-natural properties like qualia. The scientific dualist is also committed to naturalism, but takes it that the current inventory of nature is simply incomplete and that there must be a new, exotic fundamental property in the universe, that of phenomenality.

The Relativistic Approach

A common thread connecting both extremes of dualism and illusionism is that both assume that phenomenal consciousness is an absolute phenomenon, wherein an object O evinces either property P or $\neg P$. We will show that we need to abandon this assumption. The relativistic principle in modern physics posits a universe in which for many properties an object O evinces either property P or $\neg P$ *with respect to some observer X* . In such a situation, there is no one answer to the question of whether object O has property P or not. We propose a novel relativistic theory of consciousness in which consciousness is not an absolute property but a relative one. This approach eschews both extremes of illusionism and dualism. The relativistic theory of consciousness will show that phenomenal consciousness is neither an illusion created by a “machine stuck in a logic loop” nor a unique fundamental property of the universe. It will give a coherent answer to the question of the (supposed) privacy of phenomenal consciousness, will bridge the explanatory gap, and will provide a solution to the hard problem based in relativistic physics. General notions of this approach can be found in some dual aspect monisms such as Max Velmans’ reflexive monism. According to [Velmans \(2009, p. 298\)](#), “[i]ndividual conscious representations are perspectival.” Here, however, we develop a physical theory of consciousness as a relativistic phenomenon and formalize the perspectival relations in light of the relativistic principle. To do that, in the following section, we will develop more formally the *relativistic principle* and introduce the *equivalence principle of consciousness*.

In physics, relativity means that different observers from different frames of reference will nevertheless measure the same laws of nature. If, for example, one observer is in a closed room in a building and the other observer is in a closed room in a ship (one moving smoothly enough on calm water), then the observer in the ship would not be able to tell whether the ship is moving or stationary. Each will obtain the same results for any experiment that tries to determine whether they are moving or not. For both of them, the laws of nature will be the same, and each will conclude that they are stationary. For example, if they throw a ball toward the room’s ceiling, each will determine that the ball will return directly into their hands (because the ship moves with constant velocity and because of Newton’s First Law, the ball will preserve the velocity of the ship while in the air, and will propagate forward with the same pace as the ship. As a result, it will fall directly into the observer’s hands). There will be no difference in the results of each observer’s measurements, trying to establish whether they are stationary or not. They will conclude that they have the same laws of nature currently in force, causing the same results. These results will be the results of a stationary observer and thus both of them will conclude that they are at rest. Because each of them will conclude that they are the stationary one, they will not agree about one another’s status. Each of them will conclude that the other is the one that moves (common sense will tell the observer in the ship that they are the one moving, but imagine an observer locked on the ship in a room with no windows. Such an observer cannot observe the outside world. This kind of observers will conclude that they are stationary because velocity is relativistic).

To state that consciousness is a relativistic phenomenon is to state that there are observers in different *cognitive frames of reference*, yet they will nevertheless measure the same laws of nature currently in force and the same phenomenon of consciousness

in their different frames. We will start with an equivalence principle between a conscious agent, like a human being, and a zombie agent, like an advanced artificial cognitive system. As a result of this equivalence, we will show that if the relativistic principle is true, then zombies are not possible. Instead, every purported zombie will actually have phenomenal consciousness and any system with adequate functional consciousness will exhibit phenomenal consciousness from the first-person cognitive frame of reference. Others have similarly claimed that zombies are physically impossible ([Brown, 2010](#); [Dennett, 1995](#); [Frankish, 2007](#); [Nagel, 2012](#)), but our aim here is to show why that is according to the relativistic principle. As a result of this equivalence, observations of consciousness fundamentally depend on the observer's cognitive frame of reference. The first-person cognitive frame of reference is the perspective of the cognitive system itself ([Solms, 2021](#)). The third-person cognitive frame of reference is the perspective of any external observer of that cognitive system. Phenomenal consciousness is only seemingly private because in order to measure it one needs to be in the appropriate cognitive frame of reference. It is not a simple transformation to change from a third-person cognitive frame of reference to the first-person frame, but in principle it can be done, and hence phenomenal consciousness isn't private anymore. We avoid the term "first-person perspective" because of its occasional association with immaterial views of consciousness; cognitive frames of reference refer to physical systems capable of representing and manipulating inputs. These systems have physical positions in space and time and instantiate distinct dynamics.

In section "The Equivalence Principle of Consciousness: Mathematical Description," we will show that from its own first-person cognitive frame of reference, the observer will observe phenomenal consciousness, but any other observer in a third-person cognitive frame of reference will observe only the physical substrates that underlie qualia and eidetic structures. The illusionist mistake is to argue that the third-person cognitive frame of reference is *the* proper perspective. To be clear, the first-person cognitive frame of reference is still a physical location in space and time (it is not immaterial); it is just the position and the dynamics of the cognitive system itself. As we will see, this is the position from which phenomenal consciousness can be observed. The principle of relativity tells us that there is no privileged perspective in the universe. Rather, we will get different measurements depending on the observer's position. Since consciousness is relativistic, we get different measurements depending on whether the observer occupies or is external to the cognitive system in question. Both the first-person and third-person cognitive frames of reference describe the same reality from two different points of view, and we cannot prefer one point of view upon the other ([Solms, 2021](#)). The dualist mistake is to argue that phenomenal consciousness is private. For any relativistic phenomenon there is a formal transformation between the observers of different frames of reference, meaning that both frames can be accessible to every observer with the right transformations. Consciousness as a relativistic phenomenon also has such transformation rules. We will describe the transformations between first-person (i.e., phenomenological) and third-person (i.e., laboratory point of view) cognitive frames of reference. There are several consequences of these transformations. First, qualia and eidetic structures are not private. Rather, they only *appear* private, because in order to measure them one needs to be in the appropriate cognitive frame of reference, i.e., within the perspective of the cognitive system in question. We can use

these transformations to answer questions like, “What is it like to be someone else?” Because of the transformations, results that we obtain from third-person methodology should be isomorphic to first-person structures. Isomorphism between two elements means that they have the same mathematical form and there is a transformation between them that preserves this form. Equality is when two objects are exactly the same, and everything that is true about one object is also true about the other. However, an isomorphism implies that everything that is true about *some* properties of one object’s structure is also true about the other. In section “The Equivalence Principle of Consciousness: Mathematical Description,” we will show that this is the case with measurements obtained from first-person and third-person frames of reference. We will show that this isomorphism is a direct result of the relativistic principle and the notion that phenomenal judgements and phenomenal structures are two sides of the same underlying reality. All that separates them are different kinds of measurements (causing different kinds of properties). An unintuitive consequence of the relativistic theory is that the opposite is also true, and first-person structures also bear formal equivalence to third-person structures. We advocate for interdisciplinary work between philosophers and cognitive neuroscientists in exploring this consequence.

The Equivalence Principle of Consciousness

The Principle of Relativity and the Equivalence Principle

Our task is to establish the equivalence principle of consciousness, namely, that qualitative and quantitative aspects of consciousness are formally equivalent. We start by establishing the equivalence between conscious humans and zombies, and then we expand that equivalence to all structures of functional consciousness. We begin with a philosophical defense of the equivalence principle and then develop a mathematical formalization. We must first present the principle of relativity and the equivalence principle in physics. Later, we will use these examples to develop a new equivalence principle and a new transformation for consciousness. To be more formal than earlier, the principle of relativity is the requirement that the equations describing the laws of physics have the same form in all admissible frames of reference ([Møller, 1952](#)). In physics, several relativistic phenomena are well-known, such as velocity and time, and the equivalence principle between uniformly accelerated system and a system under a uniform gravitational field. Let’s examine two examples with the help of two observers, Alice and Bob.

In the first example of a simple Galilean transformation, Alice is standing on a train platform and measures the velocity of Bob, who is standing inside a moving train. Meanwhile, Bob simultaneously measures his own velocity. As the train moves with constant velocity, we know that the laws of nature are the same for both Alice’s and Bob’s frames of reference (and thus that the equations describing the laws of physics have the same form in both frames of reference; [Einstein et al., 1923/1952](#), p. 111). According to his measurements, Bob will conclude that he’s stationary, and that Alice and her platform are moving. However, Alice will respond that Bob is mistaken, and that *she* is stationary while Bob and the train are moving. At some point, Alice might say

to Bob that he has an illusion that he is stationary and that she is moving. After all, it is not commonsensical that she, along with the platform and the whole world, are moving. Still, although it doesn't seem to make common sense, in terms of physics all of Bob's measurements will be consistent with him being stationary and Alice being the one who moves. In a relativistic universe, we cannot determine who is moving and who is stationary because all experiment results are the same whether the system is moving with constant velocity or at rest. Befuddled, Bob might create an elaborate argument to the effect that his measurement of being stationary is a private measurement that Alice just can't observe. But, of course, both of their measurements are correct. The answer depends on the frame of reference of the observer. Yet, both of them draw mistaken conclusions from their correct measurements. Bob has no illusions, and his measurements are not private. Velocity is simply a relativistic phenomenon and there is no answer to the question of what the velocity of any given body is *without reference to some observer*. Their mistakes are derived from the incorrect assumption that velocity is an absolute phenomenon. As counterintuitive as it may seem, the relativity principle tells us that Bob is not the one who is "really" moving. Alice's perspective is not some absolute, "correct" perspective that sets the standard for measurement, although we may think that way in our commonsense folk physics ([Forshaw and Smith, 2014](#)). Alice is moving relative to Bob's perspective, and Bob is moving relative to Alice's perspective. Furthermore, the fact that Alice and Bob agree about all the results from their own measurements means that these two frames of reference are physically equivalent (i.e., they have the same laws of nature in force and cannot be distinguished by any experiment). Because of this physical equivalence, they will not agree about who is at rest and who is moving.

Later, Einstein extended the relativity principle by creating special relativity theory. The Galilean transformation showed that there is a transformation between all frames of reference that have constant velocity relative to each other (inertial frames). Einstein extended this transformation and created the Lorentz transformation, which describes more accurately the relativistic principle. enabling us to move from measurements in one inertial frame of reference to measurements in another inertial frame of reference, even if their velocity is near the speed of light ([Einstein, 1905](#); [Forshaw and Smith, 2014](#)). According to the transformation, each observer can change frames of reference to any other inertial frame by changing the velocity of the system. The transformation equation ensures that in each frame we'll get the correct values that the system will measure. For example, the observer will always measure that its own system is at rest and that it is the origin of the axes. Indeed, this is the outcome of the transformation equation for the observer's own system. Mathematically, a frame of reference consists of an abstract coordinate system, and (in tensor formulation) we denote each frame of reference by a different Greek letter (and by adding a prime symbol), usually by μ and ν , which are indexes for the elements of the vectors in each coordinate system. The Lorentz transformation is denoted by $\Lambda^{\mu\nu}$. It is a matrix that gets elements of a vector in one frame of reference (x'^{μ}) and gives back the elements of a vector in another frame of reference (x^{ν}):

$$x^{\nu} = \Lambda^{\nu}_{\mu} x'^{\mu} \quad (1)$$

This equation coheres with the relativistic principle. That is, it describes the laws of physics with the same form in all admissible frames of reference. This form will stay the same regardless of which frame we choose. It allows us to switch frames of reference and get the measured result in one frame, depending only on the other frame and the relative velocity between the two.

The second example comes from Albert [Einstein's \(1907\)](#) observation of the equivalence principle between a uniformly accelerated system (like a rocket) and a system under a uniform gravitational field (like the Earth). Here, Einstein extended the relativity principle even more, not only to constant velocity but also to acceleration and gravity (for local measurements that measure the laws of nature near the observer). He started with two different frames of reference that have the exact same results from all measurements made in their frames of reference. Then he used the relativity principle, concluding that because they cannot be distinguished by any local experiment they are equivalent and have the same laws of nature. Lastly, he concluded that because of the equivalence, we can infer that phenomena happening in one frame of reference will also happen in the other. Now, assume that Alice is skydiving and is in freefall in the Earth's atmosphere, while Bob is floating in outer space inside his spaceship. Although Alice is falling and Bob is stationary in space, both will still obtain the same results of every local experiment they might do. For example, and if they were to release a ball from their hand, both of them will measure the ball floating beside of them (in freefall, all bodies fall with the same acceleration and appear to be stationary relative to one another—this is why movies sometimes use airplanes in freefall to simulate outer space). Although they are in different physical scenarios, both will infer that they are floating at rest. Einstein concluded that because they would measure the same results, there is an equivalence between the two systems. In an equivalence state, we cannot distinguish between the systems by any measurement, and thus a system under either gravity or acceleration will have the same laws of nature in force, described by the same equations regardless of which frame it is.

Because of this equivalence, we can infer physical laws from one system to the other. For example, from knowing about the redshift effect of light in accelerated systems with no gravitational force, Einstein predicted that there should be also a redshift effect of light in the presence of a gravitational field as if it were an accelerated system. This phenomenon was later confirmed ([Pound and Rebka, 1960](#)). According to [Einstein \(1911\)](#),

“By assuming this [the equivalence principle] to be so, we arrive at a principle which...has great heuristic importance. For by theoretical consideration of processes which take place relative to a system of reference with uniform acceleration, we obtain information as to the behavior of processes in a homogeneous gravitational field” (p. 899).

Next, we argue that phenomenal consciousness is a relativistic physical phenomenon just like velocity. This allows us to dissolve the hard problem by letting consciousness be relativistic instead of absolute. Moreover, we develop a similar transformation for phenomenal properties between first-person and third-person perspectives. The

transformation describes phenomenal consciousness with the same form in all admissible cognitive frames of reference and thus satisfying the relativistic principle.

The Equivalence Principle of Consciousness: Conceptual Argument

Before we begin the argument, it is essential to elaborate on what ‘observer’ and ‘measurement’ mean when applying relativistic physics to cognitive science. In relativity theory, an observer is a frame of reference from which a set of physical objects or events are being measured locally. In our case, let’s define a ‘*cognitive frame of reference*’. This is the perspective of a specific cognitive system from which a set of physical objects and events are being measured. Cognitive frame of reference is being determined by the dynamics of the cognitive system (for more details, see section “The Equivalence Principle of Consciousness: Mathematical Description”).

We use the term ‘measurement’ in as general a way as possible, from a physical point of view, such that a measurement can occur between two particles like an electron and a proton. Measurement is an interaction that causes a result in the world. The result is the measured property, and this measured property is new information in the system. For example, when a cognitive system measures an apple, it means that there is a physical interaction between the cognitive system and the apple. As a result, the cognitive system will recognize that this is an apple (e.g., the interaction may be via light and the result of it will be activation of retinal cells which eventually, after sufficient interactions, will lead to the recognition of the apple). In the case of cognitive systems, there are measurements of mental states. This kind of measurement means that the cognitive system interacts with a content-bearing cognitive state, like a representation, using interactions between different parts of the system. It is a strictly physical and public process (accessible for everybody with the right tools). As a result of this definition, the starting point of the argument is with measurements that are non-controversial, i.e., measurements that are public. These measurements include two types, measurements of behavioral reports and measurements of neural representations (like the phenomenal judgement-representations). For example, physical interaction of light between an apple and a cognitive system causes, in the end of a long process of interactions, the activation of phenomenal judgement-representation of an apple and possibly even a behavioral report (“I see an apple”).

Let us start from the naturalistic assumption that phenomenal consciousness should have some kind of physical explanation. The physicalism we assume includes matter, energy, forces, fields, space, time, and so forth, and might include new elements still undiscovered by physics. Panpsychism, naturalistic dualism, and illusionism all fall under such a broad physicalism. (It might be, for example, that physics hasn’t discovered yet that there is a basic private phenomenal element in addition to the observed known elements. If this element exists, it needs to be part of the broad physicalism.) In addition, let’s assume that the principle of relativity holds, i.e., all physical laws in force should be the same in different frames of reference, provided these frames of references agree about all the results of their measurements. Since we have accepted that consciousness is physical (in the broad sense), we can obtain a new equivalence principle for consciousness. Let’s assume two agents—Alice, a

conscious human being—and a zombie in the form of a complex, artificial cognitive system that delusionally claims to have phenomenal consciousness. Let's call this artificially intelligent zombie "Artificial Learning Intelligent Conscious Entity," or ALICE. ALICE is a very sophisticated AI. It has the capacity to receive inputs from the environment, learn, represent, store and retrieve representations, focus on relevant information, and integrate information in such a manner that it can use representations to achieve human-like cognitive capabilities.

ALICE has direct access to perceptual information and to some of its own cognitive contents, a developed concept of self, the capacity to reflect by creating representations of its internal processes and higher-order representations, and can create outputs and behavioral responses (it has a language system and the ability to communicate). In fact, it was created to emulate Alice. It has the same representations, memories, and dynamics as Alice's cognitive system and as a result, it has the same behavioral responses as Alice. But ALICE is a zombie (we assume) and doesn't have phenomenal consciousness. On the other hand, conscious Alice will agree with all of zombie ALICE's phenomenal judgments. After sufficient time to practice, ALICE will be able to produce phenomenal judgments and reports nearly identical to those of Alice. After enough representational manipulation, ALICE can say, e.g., "I see a fresh little madeleine, it looks good and now I want to eat it because it makes me happy." It can also reflect about the experience it just had and might say something like, "I just had a tasty madeleine cake. It reminded me of my childhood, like Proust. There's nothing more I can add to describe the taste, it's ineffable."

Alice and ALICE will agree about all measurements and observations they can perform, whether it's a measurement of their behavior and verbalizations, or even an "inner" observation about their own judgments of their experience, thoughts and feelings. They will not find any measurement that differs between them, although Alice is conscious, and ALICE is not. For example, they can use a Boolean operation (with yes/no output) to compare their phenomenal judgments (e.g., do both agree that they see a madeleine and that it's tasty?). Now, let's follow in Einstein's footsteps concerning the equivalence principle between a uniformly accelerated system and a system under a uniform gravitational field. Alice and ALICE are two different observers, and the fact that they obtain the same measurements agrees with the conditions of the relativity principle. Accordingly, because both of them completely agree upon all measurements, they are governed by the same physical laws. More precisely, these two observers are cognitive systems, each with its own cognitive frame of reference. For the same input, both frames of reference agree about the outcome of all their measurements, and thus both currently have the same physical laws in force. As a result, these two systems are equivalent to each other in all physical aspects and we can infer physical laws from one system to the other. According to the naturalistic assumption, phenomenal consciousness is part of physics, so the equivalence between the systems applies also to phenomenal structures. Consequently, we can infer that if Alice has phenomenal properties, then ALICE also must have them. Both Alice and ALICE must have phenomenal consciousness! ALICE cannot be a zombie, like we initially assumed, because their systems are physically equivalent. This equivalence makes it impossible for us to speak of the existence of absolute phenomenal properties in the human frame

of reference, just as the theory of relativity forbids us to talk of the absolute velocity in a system. For, by knowing that there is phenomenal consciousness in the human frame of reference and by using the broad physicalist premise, we conclude that there are physical laws that enable phenomenal consciousness in the human frame of reference. Because of the equivalence principle, we can infer that the same physical laws will be present also in the supposed “zombie” cognitive system’s frame of reference. The conclusion is that if there is phenomenal consciousness in the human frame of reference, then the “zombie” cognitive system’s frame of reference must also harbor phenomenal consciousness.

We started from the premise that ALICE is a zombie and concluded that it must have phenomenal consciousness. One of our premises is wrong: either the broad physicalism, the relativistic principle, or the existence of zombies. Most likely the latter is the odd man out, because we can explain these supposed “zombies” using a relativistic, physicalist framework. As a result, although we started from a very broad notion of physicalism and an assumption that the human has phenomenal consciousness and the “zombie” cognitive system does not, the relativity principle forces us to treat phenomenal structures as relativistic. According to the relativity principle, there is no absolute frame of reference; there are only different observers that obtain different measurements. If the observers obtain the same measurements, there cannot be anything else that influences them. There is nothing over and above the observers (no God’s-eye-view), and if they observe that they have phenomenal properties, then they have phenomenal consciousness. Because there is nothing over and above the observer, we can generalize this result even further for every cognitive system that has phenomenal judgments: Any two arbitrary cognitive observers that create phenomenal judgment-representations also have phenomenal consciousness. Zombies cannot exist (assuming their cognitive systems create phenomenal judgments). In other words, we obtain an equivalence between functional consciousness (which creates phenomenal judgments) and phenomenal consciousness. Notice that even if we start from the naturalistic dualism of Chalmers and assume a broader physics including phenomenal elements alongside other aspects in the universe, the relativistic principle still forces this kind of broad physics to have the same consequences, viz., that zombies cannot exist and that there is an equivalence between phenomenal judgement-representations and phenomenal properties. (Formally, phenomenal judgement-representations and phenomenal properties are isomorphic. They have the same mathematical form and there is a transformation between them that preserves this form. See section “The Equivalence Principle of Consciousness: Mathematical Description”). As a result, the relativistic principle undermines the dualist approach altogether.

In the next section, we develop the mathematical proof of the argument. The mathematical description elaborates the fine details of the theory and reveals new insights. Notice that a mathematical background is not necessary to understand this section, as every step includes a comprehensive explanation.

The Equivalence Principle of Consciousness: Mathematical Description

In this section we develop a mathematical description of consciousness as a relativistic phenomenon. In the beginning of the argument, we apply the relativistic principle only to publicly-observable measurements. Then, we show that we can also apply the relativistic principle to phenomenal properties and structures. To that end, we use two different arguments. Then, using the relativistic principle, we prove the equivalence principle between a conscious human agent and a purported “zombie.” Then, we expand that equivalence to all cognitive frames of reference having functional consciousness (i.e., the cause of phenomenal judgments). As a result, we prove that phenomenal judgements are equal to phenomenal properties in the cognitive frame of reference that generates them. Then, we describe the difference between the first-person and third-person perspectives and develop a transformation between them and between measurements of any cognitive frames of reference that have phenomenal consciousness. We show that this transformation preserves the form of the equation regardless of which frame we choose, and thus satisfies the relativity principle. That is, it describes the laws of physics with the same form in all admissible frames of reference.

The Three-Tier Information Processing Model for Cognitive Systems

Let’s once more assume two agents, Alice, a conscious human agent, and ALICE, a “zombie” artificial cognitive system (at least, supposedly) with phenomenal judgments. These systems parallel Chalmers’ thought experiment: ALICE is the cognitive duplicate of Alice, but lacks (we suppose) phenomenal consciousness. In order to develop a fully mathematical description, we need to know the exact equations of the cognitive systems in question. Unfortunately, to date there is no complete mathematical description of a human cognitive system, let alone an artificial one that can mimic human cognition. Instead, we use a general, simplified version, a mathematical toy model, to describe a cognitive system that creates phenomenal judgments. We use a three-tier information processing model that divides a cognitive system (S) into three parts: sensation (T), perception (P), and cognition ([Dretske, 1978, 2003](#); [Pageler, 2011](#)). Sensation picks up information from the world and transforms it, *via* transduction, into neural signals for the brain. Later, perceptual states are constructed *via* coding and representations until a percept is created. The point at which cognitive processing is thought to occur is when a percept is made available to certain operations, such as recognition, recall, learning, or rational inference ([Kanizsa, 1985](#); [Pylyshyn, 2003](#)). Cognition has many different operations, but our main interest is the module that specializes in functional consciousness (and that creates phenomenal judgments), which we label C . After recognition, this functional consciousness module needs to integrate the input and emotional information, the state of the system, the reaction of the system to the input, and self-related information in order to create representations of phenomenal judgments. We can summarize this model with the following equation:

$$S = (T, P, C) \quad (2)$$

where T is the sensation subsystem, P is the perception subsystem, C is the functional consciousness subsystem of cognition, which is ultimately responsible for phenomenal judgments (we’re not considering other cognitive subsystems here). Finally, S is the

entire cognitive system as a whole. These subsystems are a 3-tuple that form the cognitive system in a sequential manner. From T to P and to C . The three-tier model yields:

$$\vec{t}_{x_t}^i = T(\vec{x}_t) \quad (3)$$

$$\vec{p}_{x_t} = P(\vec{t}_{x_t}^i) \quad (4)$$

$$\vec{c}_{x_t} = C(\vec{p}_{x_t}) \rightarrow \vec{c}_{x_t} = S(\vec{x}_t) \quad (5)$$

where $\vec{x}_t$ is an input from the physical space outside the cognitive system that interacts with the cognitive system at time t . The input has physical properties, y_j (E.g., temperature, position, velocity, etc. $j = 1, 2, 3, \dots$ is an index of the properties). These physical properties can be detected by the sensors of the sensation subsystem (e.g., the input can be an apple and the cognitive system interacts with it by a beam of light that reached its retina at time t . Using photoreceptors in the retina, the sensation subsystem can detect physical properties of the apple like its color.) T is the sensation

subsystem, and $\vec{t}_{x_t}^i$ are the outputs of the subsystem, $i = 1, 2, 3, \dots, n$ is an index for each sensory module (i.e., $\vec{t}_{x_t}^1$ is the output vector of the sensation subsystem for visual

information, $\vec{t}_{x_t}^2$ for auditory information, $\vec{t}_{x_t}^3$ for interoception information etc.). Using the sensation subsystem, the cognitive system picks up information from the world and transforms it into neural patterns. Each pattern is a different possible state in the state space of the cognitive system. In this state space every degree of freedom (variable) of the dynamic of the system is represented as an axis of a multidimensional space and each point is a different state (e.g., the axes can represent firing rate, membrane potential, position, etc.). After a learning period a subset of these states is formed and being used by the sensation subsystem to encode inputs. This is the state space of the sensation subsystem, and it is part of the state space of the whole cognitive

system. $\vec{t}_{x_t}^i$ are vectors that point to such states in the state space of the sensation

subsystem. There is a correspondence between the states $\vec{t}_{x_t}^i$ and the captured physical properties y_j (e.g., correspondence between the color of an apple and the neural pattern that it causes in the system. light from a red apple activates photoreceptors which create a neural pattern. In this pattern, cones sensitive to red frequencies will fire

more.) P is the perception subsystem. It gets $\vec{t}_{x_t}^i$ as inputs from the sensation subsystem and returns output $\vec{p}_{x_t}$, the percept of the input $\vec{x}_t$. The perception subsystem creates representations. These are gists of important information about the input $\vec{x}_t$ that the cognitive system can use even when the input is absent. the

percept $\vec{p}_{x_t}$ is a representation unifying information about the current input from all sensory modules. To that end, the perception subsystem uses yet another subset of states from the state space. This is the state space of the perception subsystem, and it is part of the state space of the whole cognitive system. $\vec{p}_{x_t}$ is a vector that points to such a state in the state space of the perception subsystem. There is a correspondence between the states and the operations that can act on these states and between the captured physical properties y_j (e.g., one operation on states can be addition of two states to get a third state in the state space of the perception subsystem. In our toy model, this operation can correspond to integration of different physical properties. A red round shaped object, for example, can be a state that is a summation of the red and round states in the state space of the perception subsystem.) C is the functional consciousness subsystem that creates phenomenal judgments (representations that are not phenomenal states themselves, but they are about phenomenology. see introduction for details). It gets percepts as inputs from the perception subsystem and returns phenomenal judgment representation, $\vec{c}_{x_t}$ as output. As before, the functional consciousness subsystem uses yet another subset of states from the state space. $\vec{c}_{x_t}$ is a vector that points to such a state in the state space of the functional consciousness subsystem. There is a correspondence between the operations that can act on these states in the functional consciousness subsystem state space and between the captured physical properties y_j . Combining these equations with eq. 2, the conclusion is that the cognitive system S , which uses the sensation, perception and functional consciousness subsystem sequentially, has a state space that is a combination of the three subsystems state spaces. It gets as input, $\vec{x}_t$ from the physical space, and returns as output, $\vec{c}_{x_t}$, the phenomenal judgment representation, from the state space of the cognitive system. To sum, when Alice sees a red apple, for example, her perception subsystem creates a representation of sensory information, a unification of representations for “red,” “round,” and possibly a smell and texture. This integrated representation is her current percept. Later, the functional consciousness subsystem will recognize the object as an “apple” and will create an integrated representation of all relevant information about this red apple. This phenomenal judgment can be, for example, “good-looking, red, and roundish apple,” which can later be reported. Such a cognitive system S needs to integrate information from previous layers and from different modules. It needs an attention module (W) that can change the weights of representations and filter information to focus on relevant information. It needs a long-term memory module (M) to store and retrieve representations. It also needs an emotion evaluation module (E), which assesses information and creates preferences, rewards, and avoidances (e.g., does the input have a positive or negative affective valence?). Other modules could include an affordance module (A) that detects possibilities for action in the environment according to the input, and a self-module (I) that collects self-related information to create a unified representation of the self. Lastly, the cognitive system needs modules to create outputs and behavioral responses. In particular, it has a language module (L) that creates lexical, syntactic, and semantic representations and responses that the functional consciousness subsystem (C) can use. Adding them to Equations 4, 5, we obtain:

$$\vec{p}_{x_t} = P(\vec{t}_{x_t}^i, E, W, M) \quad (6)$$

the perception subsystem P receives as variables the outputs of the sensation subsystem $\vec{t}_{x_t}^i$ and uses the long-term memory module M , the attention module W , and the evaluation module E to create a percept $\vec{p}_{x_t}$ of the input $\vec{x}_t$, a unified representation of perception from all sensory modules. This yields:

$$\vec{c}_{x_t} = C(\vec{p}_{x_t}, E, W, M, A, I) \quad (7)$$

the functional consciousness subsystem C receives as a variable the output of the perception subsystem $\vec{p}_{x_t}$ and uses the long-term memory module M , the attention module W , the evaluation module E , the affordance module A and the self-module I to create a phenomenal judgment $\vec{c}_{x_t}$ (a complex representation that carries content about phenomenology) concerning the input $\vec{x}_t$. We can also add the language module for the cognitive system, so that S can understand and answer questions:

$$\vec{p}_l = L(\vec{a}_l, E, W, M) \quad (8)$$

where L is the language module, $\vec{a}_l$ is the input of the language module, a low-level processed representation from the perception subsystem (without loss of generality, let us assume it to be auditory information. It can also be a low-level representation of other sensory modules). The language module L gets a low-level representation from the perception subsystem $\vec{a}_l$ and uses the long-term memory module M , the attention module W , and the evaluation module E to create several representations. Semantic representations for words, syntactic representations, and eventually sentence comprehension representation, $\vec{p}_l$. As before, the language module using yet another subset of states from the state space. $\vec{p}_l$ is a vector that points to such a state in the state space of the language module. The functional consciousness subsystem C uses the output of the language module to perform its tasks and create phenomenal judgments (answering a question about a red apple, for example):

$$\vec{c}_{p_l, p_{x_t}} = C(\vec{p}_l, \vec{p}_{x_t}, E, W, M, A, I) \quad (9)$$

Now we add $\vec{p}_l$ to Equation 7, which is the sentence comprehension representation (a question, for example) as another variable of C . The answer, $\vec{c}_{p_l, p_{x_t}}$, is sent to the language module, this time creating a linguistic response:

$$\vec{l}_{p_l, x_t} = L(\vec{c}_{p_l, p_{x_t}}, E, W, M) \quad (10)$$

where L is the language module and $\vec{l}_{p_{l,x_t}}$ is a vector of the linguistic response. As before, the language module uses yet another subset of states from the state space. $\vec{l}_{p_{l,x_t}}$ is a vector that points to such a state in the state space of the language module. This state is a complex neural pattern caused by motor neurons that innervate muscle fibers. The pattern causes contraction patterns in the muscles to create a linguistic response. The response is according to $\vec{c}_{p_{l,p_{x_t}}}$, the phenomenal judgement representation that captures both the question and the answer of C . For example, suppose that Alice asks ALICE what she sees, and ALICE sees a red apple. The visual information will be transduced and processed in the perception subsystem until a percept is created (equation 6). The language system creates a representation of the question, $\vec{p}_l$ = “what do you see?” (eq. 8) and the functional consciousness subsystem uses the representation of the question, $\vec{p}_l$ and the percept of the apple, $\vec{p}_{x_t}$, as variables to build a phenomenal judgement to answer the question, $\vec{c}_{p_{l,p_{x_t}}}$ (eq. 9). Finally, the language module uses this representation to produce a proper linguistic response, $\vec{l}_{p_{l,x_t}}$ = “I see an apple” (eq. 10).

The Equivalence Principle of Consciousness

In parallel with Chalmers' (1996) assumptions about zombies, Alice and ALICE have equivalent cognitive systems. But Alice, being human, also has phenomenal consciousness (without loss of generality, let's assume Alice has a quale Q_{x_t} about input $\vec{x}_t$. We will not assume any inner structure for the quale). After we establish the equations for the cognitive system, we can use them to prove the equivalence principle for consciousness. Let's start with general considerations about cognitive frames of reference. A cognitive frame of reference is determined by the dynamics of a cognitive system (represented by Equations 1–10). Because the equations create and use representations, the representations are an inseparable part of the dynamics of a cognitive system. If two cognitive systems have different dynamics and representations, they are in different cognitive frames of reference. But if they have the same dynamics and representations, then they are in the same cognitive frame of reference. We will denote cognitive frames of reference by μ , and ν .

We started with the assumption that ALICE is a zombie. Hence, ALICE will obtain the same behavioral reports (i.e., $\vec{l}_{p_{l,x_t}}$) and the same phenomenal judgments ($\vec{c}_{x_t}$) as Alice, who also has phenomenal consciousness (note that for this claim to be true, we idealize the situation as if Alice and ALICE were standing in the same spatial position. Hence, they will obtain the same phenomenal judgments. We can neglect minor differences arising from their different spatial positions, as long as they are standing near each other). In fact, Alice and ALICE have the same exact cognitive structure and dynamics (i.e., the same subsystems, modules, dynamics, and representations). As a result, they

are in the same cognitive frame of reference and for each question they will have the same responses and the same phenomenal judgments. We can formalize it as such:

$$\langle Alice \rangle^v = \tilde{\Lambda}^{v\mu} \langle ALICE \rangle^\mu \quad (11)$$

where v is Alice's cognitive frame of reference, μ is ALICE's cognitive frame of reference, and $\langle \rangle$ is a symbol for the set of all results of measurements of a system (i.e., $\langle Alice \rangle$ means the set of results of measurements conducted by Alice). $\tilde{\Lambda}^{v\mu}$ is a general transformation from the set of results of measurements of frame μ to the corresponding set of results of measurements of frame v . Alice and ALICE can measure or observe both their behavioral reports and their phenomenal judgments, and then compare the results. They will discover that their results are the same. For the same input, they will not find any measurement about themselves to distinguish between them (again, the fact that they stand in slightly different spatial positions can be neglected because any reasonable cognitive system takes different positions into account in order to build percepts correctly, regardless of the system's position. For example, the cognitive system needs to recognize an apple regardless of the angle between the apple and the system. Objects should be invariant structures across changing conditions.) According to the relativistic principle, because they obtain the same results about themselves, they have the same physical laws currently in force in their respective frames of reference. There is an equivalence between the two systems, and we can infer physical laws and phenomena from one system to the other. For example, we know that

$$\cup \tilde{l}_{p_l, x_t}^v = \tilde{\Lambda}^{v\mu} \cup \tilde{l}_{p_l, x_t}^\mu \quad (12)$$

There is a transformation from the unification of all linguistic responses of Alice, $\cup \tilde{l}_{p_l, x_t}^v$, to the unification of all linguistic responses of ALICE, $\cup \tilde{l}_{p_l, x_t}^\mu$ (provided that both were asked the same questions, $\tilde{p}_l$). Let's consider a specific linguistic response. From Equation 11 we obtain:

$$\tilde{l}_{p_l, x_t}^v = \tilde{\Lambda}^{v\mu} m^\mu, \quad (13)$$

where m^μ is a specific measurement result of ALICE, $m^\mu \in \langle ALICE \rangle$. $\tilde{l}_{p_l, x_t}^v$ is a specific linguistic response of Alice. There is a transformation from the linguistic response of Alice that she will measure (i.e., that she will observe about herself) in her cognitive frame of reference to a specific measurement that ALICE will obtain about herself in her frame of reference. Now we can ask, what is the specific measurement in ALICE's frame of reference? Because of the relativistic principle we know that the frames of reference of Alice and ALICE are equivalent, and hence we just need to replace v with μ :

$$\vec{l}_{p_{l,x_t}}^v = \delta^{v\mu} m^\mu \quad (14)$$

$$\vec{l}_{p_{l,x_t}}^\mu = m^\mu \quad (15)$$

We use a special function known as the delta function, $\delta^{v\mu}$, to transform from frame v to frame μ . This function equals 0 if $v \neq \mu$, and equals 1 if $v = \mu$. In other words, the only solution that is different from 0 for Equation 14 is when $v = \mu$. Now we can plug in 1 instead of $\delta^{v\mu}$ and substitute μ for v on the left-hand side of Equation 15, with the result

that ALICE will measure $\vec{l}_{p_{l,x_t}}^\mu$. Now we can plug the result of Equation 15 back in Equation 13, which gives us a similar equation to Equation 12 for a specific linguistic response. If we repeat this process for every linguistic response of the two systems, we obtain Equation 12 (for details, see [Supplementary Note 1](#)). Proceeding in this manner allows us to infer physical laws and new phenomena from an observer in one frame of reference to the observer in the other frame. Next, we will use this process to infer the existence of qualia in ALICE's frame of reference (similar to what Einstein did with the equivalence principle between uniform acceleration and uniform gravitational field). The transformation term $\tilde{\Lambda}^{v\mu}$, is equal to the delta function because of the relativistic principle. Because Alice and ALICE obtain the same measurement results, they are in the same cognitive frame of reference.

So far, we have applied the relativistic principle only to public measurements like reports and phenomenal judgements. But we know that Alice also observes quale Q_{x_t} in her frame of reference. According to broad physicalism, phenomenal consciousness has some kind of physical explanation (physical laws that govern phenomenal consciousness) and as such it is part of physics and part of the physical measurements that Alice can conduct in her first-person frame of reference, v (even if it's a unique, private measurement). But for the relativistic principle to hold true, all physical laws currently in force in both respective frames of reference should be the same (including the physical laws that enable qualia). Thus, because of our broad physicalist assumption, we can obtain from Equation 11:

$$Q_{x_t}^v = \tilde{\Lambda}^{v\mu} \tilde{m}^\mu \quad (16)$$

where $\tilde{m}^\mu$ is yet another specific measurement result of ALICE and $Q_{x_t}^v$ is Alice's quale about input x_t . Just as before, there is a general transformation between the quale that Alice measures in her frame of reference to a specific measurement that ALICE observes about herself in her own frame of reference. Because of the relativistic principle and the equivalence between the two frames of reference, all we need to do is replace v with μ :

$$Q_{x_t}^v = \delta^{v\mu} \tilde{m}^\mu \quad (17)$$

$$Q_{x_t}^\mu = \tilde{m}^\mu. \quad (18)$$

We see that the relativistic principle yields that ALICE will also measure a quale in her frame of reference (and thus zombies cannot exist). We can plug Equation 18 back into Equation 16 and repeat this process for all Alice's qualia. As a result, Alice and ALICE will measure the same qualia about themselves (just like they obtain the same behavioral response and the same phenomenal judgments, assuming we can neglect the small differences due to their differing spatial positions).

This proof leaves the possibility that even though zombies cannot exist because of the relativistic principle, phenomenal consciousness can still be private. To accommodate that, we need to develop yet another mathematical argument, one that is more detailed (eq. 19-35). In addition, the second argument doesn't use the assumption of broad physicalism (at least not explicitly). Notice that until now we haven't used the equations of the cognitive systems. Let's go back to Equation (13), for the linguistic reports:

$$\tilde{l}_{p_l, x_t}^v = \tilde{\Lambda}^{v\mu} \tilde{l}_{p_l, x_t}^\mu \quad (19)$$

From equation (10) about the linguistic module, we obtain:

$$L\left(\tilde{c}_{p_l, p_{x_t}}, E, W, M\right)^v = \tilde{\Lambda}^{v\mu} L\left(\tilde{c}_{p_l, p_{x_t}}, E, W, M\right)^\mu. \quad (20)$$

We can obtain similar equations for the functional consciousness sub-system (from Equation (9)):

$$\begin{aligned} \tilde{c}_{p_l, p_{x_t}}^v &= C(\tilde{p}_l, \tilde{p}_{x_t}, E, W, M, A, I)^v = \\ \tilde{\Lambda}^{v\mu} C(\tilde{p}_l, \tilde{p}_{x_t}, E, W, M, A, I)^\mu &= \tilde{\Lambda}^{v\mu} \tilde{c}_{p_l, p_{x_t}}^\mu \end{aligned} \quad (21)$$

Now, consider a special case where Bob asks ALICE and Alice a question about the experience they just had (for example, $\tilde{p}_l = \text{"What are you experiencing right now?"}$). In this case, the cognitive system needs to check the phenomenal judgment representation itself. As a result, the functional consciousness subsystem will have the last phenomenal judgments as an input, $\tilde{c}_{x_t}$. Plugging it into Equation 9 gives:

$$\tilde{c}_{p_l, c_{x_t}} = C(\tilde{p}_l, \tilde{c}_{x_t}, E, W, M, A, I). \quad (22)$$

The answer, $\vec{c}_{p_l, c_{x_t}}$, is yet another phenomenal judgment representation concerning the previous phenomenal judgment and the question. The linguistic module will use this representation to create a linguistic response according to Equation 10:

$$\vec{l}_{p_l, c_{x_t}} = L \left(\vec{c}_{p_l, c_{x_t}}, E, W, M \right). \quad (23)$$

The linguistic response can be, e.g., $\vec{l}_{p_l, c_{x_t}} = \text{"I'm experiencing happiness right now."}$ In this scenario we can write the same transformation between ALICE and Alice for the functional consciousness subsystem and the linguistic module as before:

$$\begin{aligned} \vec{c}_{p_l, c_{x_t}}^v &= C \left(\vec{p}_l, \vec{c}_{x_t}, E, W, M, A, I \right)^v \\ &= \tilde{\Lambda}^{v\mu} C \left(\vec{p}_l, \vec{c}_{x_t}, E, W, M, A, I \right)^\mu = \tilde{\Lambda}^{v\mu} \vec{c}_{p_l, c_{x_t}}^\mu \end{aligned} \quad (24)$$

$$\begin{aligned} \vec{l}_{p_l, c_{x_t}}^v &= L \left(\vec{c}_{p_l, c_{x_t}}, E, W, M \right)^v = \tilde{\Lambda}^{v\mu} \\ L \left(\vec{c}_{p_l, c_{x_t}}, E, W, M \right)^\mu &= \tilde{\Lambda}^{v\mu} \vec{l}_{p_l, c_{x_t}}^\mu \end{aligned} \quad (25)$$

For all these transformations in Equations 19–21, 24–25, we can substitute the general transformation term, $\tilde{\Lambda}^{v\mu}$, with a delta function, because of the equivalence between Alice's and ALICE's frames of reference (both have the same cognitive system and measurements):

$$\tilde{\Lambda}^{v\mu} = \delta^{v\mu} \quad (26)$$

For example, Equations 24–25 will get the form:

$$\begin{aligned}\vec{c}_{p_{l,c_{x_t}}}^v &= C(\vec{p}_l, \vec{c}_{x_t}, E, W, M, A, I)^v = \\ \delta^{v\mu} C(\vec{p}_l, \vec{c}_{x_t}, E, W, M, A, I)^\mu &= \delta^{v\mu} \vec{c}_{p_{l,c_{x_t}}}^\mu\end{aligned}\quad (27)$$

$$\begin{aligned}\vec{l}_{p_{l,c_{x_t}}}^v &= L(\vec{c}_{p_{l,c_{x_t}}}, E, W, M)^v = \\ \delta^{v\mu} L(\vec{c}_{p_{l,c_{x_t}}}, E, W, M)^\mu &= \delta^{v\mu} \vec{l}_{p_{l,c_{x_t}}}^\mu\end{aligned}\quad (28)$$

Now, let's switch to Alice's frame of reference. From Alice's point of view, she has phenomenal consciousness. This perspective is also a physical frame of reference. From her first-person perspective, she experiences happiness (quale Q_{x_t}) and a question, Q_l = "What are you experiencing right now?" After the question, she will have an experience of an answer to the question, $Q_{Q_l, Q_{x_t}}$, and accordingly she will respond with $\vec{l}_{Q_l, Q_{x_t}}$ = "I'm experiencing happiness right now." Notice that from her first-person frame of reference, she directly experiences her qualia and uses the quale (e.g., of happiness) to answer the question (in her head) and to articulate the answer as a linguistic report. We can formulate an equation according to her first-person frame of reference:

$$\vec{l}_{Q_l, Q_{x_t}}^v = F(Q_{Q_l, Q_{x_t}})^v, \quad (29)$$

where $Q_{Q_l, Q_{x_t}}$ is the quale of the answer, $\vec{l}_{Q_l, Q_{x_t}}$ is the linguistic response, and F is a general function mapping a quale to a vector of the linguistic response. From Alice's frame of reference, she experiences herself trying to answer the question in her head ($Q_{Q_l, Q_{x_t}}$) and then responds accordingly ($\vec{l}_{Q_l, Q_{x_t}}$). Her direct experience is that her linguistic response, $\vec{l}_{Q_l, Q_{x_t}}$ is a result of the quale of her answer, $Q_{Q_l, Q_{x_t}}$.

We can now make use of the equivalence between Alice and ALICE. First, notice that we can describe Alice's linguistic response according to her first-person perspective (Equation 29) and according to a third-person perspective (Equation 23). Both

descriptions refer to the same linguistic response, and thus $\vec{l}_{Q_l, Q_{x_t}}^v = \vec{l}_{p_{l,c_{x_t}}}^v$. Plugging this result into Equation 29 gives us:

$$\vec{l}_{p_{l,c_{x_t}}}^v = F(Q_{Q_l, Q_{x_t}})^v. \quad (30)$$

According to Equation 28, because of the equivalence between Alice's and ALICE's frames of reference for linguistic responses, we can replace the left-hand side of Equation 30 with:

$$\delta^{v\mu} \bar{l}_{p_{l,c_{x_t}}}^v = F(Q_{Q_l}, Q_{x_t})^v. \quad (31)$$

Because of the delta function, we know that $v = \mu$, and we can substitute μ with v on the right-hand side of Equation 31:

$$\bar{l}_{p_{l,c_{x_t}}}^v = F(Q_{Q_l}, Q_{x_t})^\mu. \quad (32)$$

We have thus completed the transformation to ALICE's frame of reference. ALICE, like Alice, must have qualia from her first-person perspective, which she uses to create a linguistic response. As a result of Equation 32, we notice again that the relativistic principle yields that ALICE, as cognitive frame of reference μ , will also observe a quale in her frame of reference. This invalidates Chalmers' (1996) assumption that a zombie can be physically and computationally isomorphic to a phenomenally conscious person (which we assumed in the beginning by stating that ALICE was a zombie). Notice that

now we can use Equation 23 and substitute $\bar{l}_{p_{l,c_{x_t}}}^\mu$ in Equation 32 with:

$$L\left(\bar{c}_{p_{l,c_{x_t}}}, E, W, M\right)^\mu = F(Q_{Q_l}, Q_{x_t})^\mu. \quad (33)$$

There is an equivalence between the form of the right-hand side of the equation and the form of the left-hand side of the equation (in both sides there is a function that gets input). We can identify the function F as the function of the linguistic module, $F = L$, and the input of function F as the input of function L (notice that the modules E, W, M are not the inputs of L , they are just modules that the dynamics of the language module uses to create the linguistic response according to the input $\bar{c}_{p_{l,c_{x_t}}}$):

$$\bar{c}_{p_{l,c_{x_t}}}^\mu = Q_{Q_l, Q_{x_t}}^\mu. \quad (34)$$

We here get an interesting result in ALICE's frame of reference. Not only will ALICE observe a quale just like Alice, but this quale is exactly her complex integrated representation that bears content about phenomenology ('phenomenal judgement'). We

can identify between the inputs $\bar{c}_{p_{l,c_{x_t}}}^\mu$ and $Q_{Q_l, Q_{x_t}}^\mu$ because in conscious agents like humans there is a close relation between the two. In fact, there is an isomorphism between them, and we can do a one-to-one correspondence mapping between them. As we saw, every quale will always be accompanied with a corresponding phenomenal judgment and every phenomenal judgment will always be accompanied with a corresponding quale. This pair of quale and phenomenal judgment is unique for each

input $\bar{x}_t$ and every time $\bar{x}_t$ is present for the conscious cognitive system, if one part of the pair is present then the other part will be present as well. We saw that zombies cannot exist because they must have qualia according to the relativistic principle (eq. 32). We also saw that in the cognitive frame of reference that uses phenomenal judgments there is an equality between the functions that use qualia and phenomenal judgments (eq. 33). The identity between the functions and the isomorphism between qualia and phenomenal judgments are enough to show that in the cognitive frame of reference that uses phenomenal judgments there is an equality between qualia and phenomenal judgments (eq. 34). We reach an equivalence between phenomenal consciousness and phenomenal judgments that were created by cognitive systems, and hence obtain an equivalence between functional consciousness and phenomenal consciousness. It is easy to show that if we use the equivalence principle once again to move back to Alice's frame of reference, we will obtain the same result for her, as well. We developed this equation from a special case of a quale about a quale (the

experience of an experience, for example, $\bar{I}_{Q_t, Q_{x_t}} = \text{"I'm experiencing happiness right now."}$), but we can repeat the same process for different cases of qualia, such as a quale of an arbitrary input (seeing an apple for example), Q_{x_t} . Generally, from the first-person perspective, a quale causes something to happen, whether it will cause a linguistic response (like in Equation 29) or the next associative thought. Because there is always an equivalent phenomenal judgment representation that will do the same thing in the cognitive system, we can always repeat a similar process like we did here (Equations 29–34) with an adequate equation similar to Equation 29 for the specific case. Furthermore, we can use the identity in Equation 34 in these similar equations to prove similar identities for different kinds of qualia (see [Supplementary Note 2](#) for details). The relativistic principle ensures that because we cannot distinguish between the quale and the phenomenal judgment representation that accompanies it, they are the same thing. Consequently, we can generalize Equation 34:

$$\bar{c}_{x_t}^\mu = \bar{Q}_{x_t}^\mu \quad (35)$$

In the first-person frame of reference, there is an equality between the quale and the phenomenal judgment-representation, and hence the quale is a point in an appropriate state space and can be represented as a vector. The first-person frame of reference is the frame of the cognitive system measuring itself. In other words, in its dynamics, the cognitive system uses its own representations as inputs and outputs and they interact with each other (like in our cognitive system model, Equations 1–10).

Conditions for Third- and First-Person Frames of Reference

What happens when we move to different cognitive frames of reference? Why can't they directly measure the quale of the other frame of reference? Why is there a difference between first-person and third-person perspectives, if it's not because of some phenomenal property of privacy? To answer these questions, let's expand the equations beyond the simple case of the equivalence between a conscious human and a copy of the human's cognitive system (a zombie-like system). We showed that in the simple case of Alice and ALICE, not only are both cognitive systems equivalent, but they are

also in the same cognitive frame of reference, $v = \mu$ (that's why we could use the delta function).

Now, let's examine two cognitive frames of reference that are not a copy of each other: Alice and Bob. These two cognitive systems developed separately and there is no reason to assume that they learned to associate inputs with the same neural patterns. Each associated different states as representations from the state space according to its own personal developmental history. As a result, Alice and Bob have no states in common and the set of states that Alice's cognitive system uses is disjoint with the set of states that Bob's cognitive system uses. As before, from her first-person frame of

reference, Alice has an experience of an apple (quale $\bar{Q}_{x_t}^v$). What will happen if Bob tries to directly observe the quale of Alice in his cognitive frame of reference? We can describe this process from his cognitive frame of reference, μ . His perception, functional consciousness, and linguistic subsystems (Equations 6–8) directly obtain Alice's quale:

$$P(\bar{Q}_{x_t}^v, E, W, M)^\mu = P(\bar{c}_{x_t}^v, E, W, M)^\mu = \emptyset \quad (36)$$

$$C(\bar{Q}_{x_t}^v, E, W, M, A, I)^\mu = C(\bar{c}_{x_t}^v, E, W, M, A, I)^\mu = \emptyset \quad (37)$$

$$L(\bar{Q}_{x_t}^v, E, W, M)^\mu = L(\bar{c}_{x_t}^v, E, W, M)^\mu = \emptyset, \quad (38)$$

where $\emptyset$ is the empty set. To measure Alice's quale directly means to use it directly in Bob's cognitive frame of reference. Because her quale is equal to a phenomenal

judgment representation (Equation 35), we substitute $\bar{Q}_{x_t}^v$ with $\bar{c}_{x_t}^v$. But because Bob and Alice are two different cognitive frames of reference, Bob's cognitive system doesn't recognize Alice's representation, and we get an empty set. As a result, Bob cannot directly measure Alice's quale. The only solution for Bob is to indirectly measure it using his sensation subsystem (Equation 3).

What will Bob measure while using his sensation subsystem to measure Alice's quale? Recall that the sensation subsystem measures physical properties, y_j from outside of the cognitive system, that can be detected by its sensors. According to eq. 35, qualia, like phenomenal judgments, are neural patterns or states of the state space of Alice's cognitive system (that are being measured directly in Alice's cognitive frame of reference). every degree of freedom (variable) of the dynamics of the system is an axis of this space (e.g., the axes can be firing rate, membrane potential, position of firing neurons, etc.). The axes form the basis vectors of the state

space, $\sum_{m=1}^N \hat{e}_m = \hat{e}_1 + \hat{e}_2 + \dots + \hat{e}_N$ where $\sum \hat{e}$ means a summation over the basis vectors and $m = 1, 2, 3, \dots, N$ is an index that runs from 1 to N, according to the N-dimensions of the state space. Each state can be written, in a unique way,

as $\sum_{m=1}^N \alpha_m \hat{e}_m$, where $\alpha_1 \dots \alpha_N$ are coefficients. These Numbers are the coordinates of the axes (e.g., suppose that for the neural representation of an apple there is a specific

firing rate. This is the coordinate α_1 of the first axis $\hat{e}_1$, the axis of firing rates and we will denote it by, $\alpha_1 \hat{e}_1$.) These variables and coefficients of the dynamics of Alice's cognitive system are the physical properties y_j that Bob's sensation subsystem will measure in the case of Alice's quale/representation, $y_j = \sum_{m=1}^j \alpha_m \hat{e}_m$ (e.g., without losing generality, let's assume that Alice's quale/phenomenal judgement while seeing an apple will have a specific pattern of 3 variable: firing rates, membrane potentials and position of firing neurons. Each of them with a specific coefficient. Bob's sensation subsystem will measure these 3 properties with their specific coefficients.):

$$T(\vec{Q}_{x_t}^v)^\mu = T(\vec{c}_{x_t}^v)^\mu = \vec{t}_{Q_{x_t}}^\mu = \vec{t}^\mu \sum \alpha_m \hat{e}_m, \quad (39)$$

where T^μ is Bob's sensation subsystem, $\vec{Q}_{x_t}^v$ is Alice's quale, $\vec{c}_{x_t}^v$ is Alice's phenomenal judgment representation, $\vec{t}_{Q_{x_t}}^\mu$ are the outputs of the sensation subsystem of Bob about Alice's quale. $\sum \alpha_m \hat{e}_m$ are the variables and coefficients of Alice's quale/representation, and $\vec{t}^\mu \sum \alpha_m \hat{e}_m$ are the outputs of the sensation subsystem of Bob about the physical properties of Alice's quale. Bob's sensation subsystem is the only part of the cognitive system that can measure Alice's quale, though indirectly. As before, we substitute her quale with the corresponding representation. The sensation subsystem can measure the physical properties of the representation and create an output that Bob's cognitive frame of reference can utilize (e.g., a visual sensation of a brain pattern). Only then will Bob's cognitive frame create its own percept, and eventually a quale of this sensation. Notice that this quale is about the indirect measurement of Alice's quale, a brain pattern that Bob measured from Alice's brain:

$$C\left(P\left(\vec{t}^\mu \sum \alpha_m \hat{e}_m\right)\right)^\mu = \vec{c}_{\sum \alpha_m \hat{e}_m}^\mu = \vec{Q}_{\sum \alpha_m \hat{e}_m}^\mu \quad (40)$$

Notice the difference between Alice's initial quale, $\vec{Q}_{x_t}^v$, and Bob's final quale about her brain pattern, $\vec{Q}_{\sum \alpha_m \hat{e}_m}^\mu$. Interestingly, even in the same cognitive frame of reference (of Alice, for example), if we introduce to Alice's sensation subsystem one of Alice's qualia, the sensation subsystem will still not recognize it directly as a quale. Instead, the subsystem will measure the physical properties corresponding to that quale, $y_j = \sum_{m=1}^j \alpha_m \hat{e}_m$. For example, we can show Alice an activation pattern going on in her brain as she reports seeing an apple. Alice will not recognize this pattern as her representation of seeing an apple and will instead see a visual image of the brain pattern. This is because the sensation subsystem still doesn't use representations, let alone the representations of higher cognitive subsystems like perception and cognition. It just measures physical properties from outside of the cognitive system:

$$T(\vec{Q}_{x_t}^v)^v = T(\vec{c}_{x_t}^v)^v = \vec{t}_{\sum \alpha_m \hat{e}_m}^v \quad (41)$$

Only the parts of the cognitive system that use the phenomenal judgment representation, $\vec{c}_{x_t}^\mu$, will measure it directly (such as the functional consciousness subsystem and the language module). The relativistic principle ensures us that this direct measurement of a phenomenal judgment representation is a quale (Equation 35). Any other cognitive frames of reference will just measure the physical substance of the representation *via* the sensation subsystem. It seems that the reason we have such a distinct difference between first-person and third-person perspectives is due to the direct measurement of the phenomenal judgment representation in the functional consciousness subsystem, and the function of the sensation subsystem that specializes in measuring bare physical properties and not representations. In sum, the condition for the ability to measure a quale (viz., the first-person frame of reference) is to have a subsystem that uses the corresponding phenomenal judgment representation (i.e., to measure directly phenomenal judgment. Because then the functional consciousness subsystem uses the phenomenal judgment in its dynamics and according to eq. 35, we can substitute it with the appropriate quale):

$$\begin{aligned} C(\vec{c}_{x_t}^\mu, E, W, M, A, I)^\mu &= C(\vec{Q}_{x_t}^\mu, E, W, M, A, I)^\mu \\ &= \vec{c}_{c_{x_t}}^\mu = \vec{Q}_{Q_{x_t}}^\mu \end{aligned} \quad (42)$$

Here, the functional consciousness subsystem, C^μ , uses in its dynamics (directly measures) a phenomenal judgment representation, $\vec{c}_{x_t}^\mu$, and hence measures it directly to create a new phenomenal judgment representation about it, $\vec{c}_{c_{x_t}}^\mu$. According to the relativistic principle, these phenomenal judgments are measured as qualia, $\vec{Q}_{x_t}^\mu, \vec{Q}_{Q_{x_t}}^\mu$. The condition to have a third-person frame of reference is to activate the sensation subsystem (39, 41):

$$T(\vec{Q}_{x_t})^\mu = \vec{t}^i{}^\mu \sum \alpha_m \hat{e}_m \quad (43)$$

Notice that the cognitive frame of reference that measured the quale is irrelevant for the sensation subsystem. The output will always be the same: sensations of the physical properties of the quale/phenomenal judgement, $\vec{t}^i{}^\mu \sum \alpha_m \hat{e}_m$.

Transformation Between Cognitive Frames of Reference and Between First- and Third-Person Frames of Reference

Finally, according to the principle of relativity, all admissible frames of reference measure the same laws of physics that are currently in force, and hence the equations describing the laws of physics have the same form in all admissible frames of reference (like Equation 1, describing the Lorentz transformation between all inertial frames of reference). As a relativistic phenomenon, consciousness should also have a similar transformation that preserves the same form of the transformation equation for all admissible cognitive frames of reference. Let's develop this equation. The transformation should take us from the measurements of one frame, μ (Bob), to the

measurements of another frame, v (Alice). Regardless of the measurements, the equation should look the same (for example, if one of the measurements causes new terms to appear in the equation, the form of the equation is no longer the same, and thus the equation is not relativistic because we can distinguish between the measurements according to their different terms. As a result, the equivalence between the admissible cognitive frames of reference has broken. We need to ensure that this is not the case in our equation). Notice that, in contrast to the scenario of Alice and ALICE, now the transformation has no constraints in the form of a question that guides the observers. Instead, they measure the same input, and the transformation should give the answer for what the cognitive system in each frame of reference measures. When we move from measurements of a quale in one cognitive frame of reference, μ , to measurements of a quale in a second cognitive frame of reference, v , there are three cases of what they can measure. The cognitive frames can measure a third input (e.g., an apple), they can measure frame μ , or they can measure frame v . If they measure a third input, like an apple, each of the frames will have its own quale about the apple, $\vec{Q}_{x_t}^\mu$, $\vec{Q}_{x_t}^v$. If they measure the quale of frame v , then frame v measures the quale $\vec{Q}_{Q_{x_t}^\mu}^\mu$ directly, while frame v will measure the physical properties of the quale of frame μ , $\vec{Q}_{\sum \alpha_m \hat{e}_m}^v$, and vice versa if they measure the quale of frame v .

We need to use another transformation term, $\Lambda^{v\mu}$, for this general scenario. It needs to give the correct measurements for each case, while preserving the form of the transformation equation. Because we lack the accurate equations of the cognitive system and use general forms in the equations, the transformation equation will also be in a general form:

$$\vec{Q}_{x_t}^v = \Lambda^{v\mu} (\vec{Q}_{x_t}^\mu), \quad (44)$$

where $\Lambda^{v\mu}$ is the transformation function from cognitive frame μ to cognitive frame v , and $\vec{Q}_{x_t}^v$, $\vec{Q}_{x_t}^\mu$ are the qualia of input x_t as measured in frames v , μ respectively.

In order to build $\Lambda^{v\mu}$, for the transformation to operate successfully, one key element is to check the physical properties, y_j of the input. To that end we introduce the function m_C that checks if the physical properties y_j of the input $\vec{x}_t$ are the physical properties of the representations that the functional consciousness subsystem uses as outputs, $\sum_{m=1}^j \alpha_m \hat{e}_m$. In other words, it checks whether the input is a state of the state space of the functional consciousness subsystem C (i.e., phenomenal judgments). If it does then m_C returns the state $\sum_{m=1}^j \alpha_m \hat{e}_m$ and if the input is not part of the state space of C , then m_C returns the empty set $\emptyset$:

$$\begin{aligned}
m_C(\vec{x}_t) &= \sum_{m=1}^j \alpha_m \hat{e}_m \leftrightarrow C(\vec{x}_t, E, W, M, A, I) \\
&= C\left(\sum_{m=1}^j \alpha_m \hat{e}_m, E, W, M, A, I\right) \\
&= C(\vec{c}_{\vec{x}_t}, E, W, M, A, I) \\
&\text{else } m_C(\vec{x}_t) = \emptyset.
\end{aligned} \tag{45}$$

If the input $\vec{x}_t$ is a state of the state space of C then it's a phenomenal judgment, $\vec{x}_t = \sum_{m=1}^j \alpha_m \hat{e}_m = \vec{c}_{\vec{x}_t}$ where $\vec{c}_{\vec{x}_t}$ is a phenomenal judgment about some input $\vec{x}_t$. In that case, m_C returns this state. Notice that because C can get phenomenal judgments as inputs (see eq. 22 and 42) $\vec{x}_t$ can be input of C and thus can be measured directly by the functional consciousness subsystem. If the input is not part of the state space of C , then m_C returns $\emptyset$. Now we can introduce the transformation function $\Lambda^{v\mu}$:

$$\Lambda^{v\mu}(\vec{x}_t, S^v) = \begin{cases} m_C^v(\vec{x}_t) = \emptyset \vee \delta^{\vec{r}_0^v \vec{r}_{x_t}^v} = 0, & S^v(\vec{x}_t) \\ m_C^v(\vec{x}_t) \neq \emptyset \wedge \delta^{\vec{r}_0^v \vec{r}_{x_t}^v} = 1, & C^v(\vec{x}_t) \end{cases} \tag{46}$$

Where v is a symbol of the logic operation 'or' and $\wedge$ is a symbol of the logic operation 'and.' In order to carry out the transformation from cognitive frame μ to cognitive frame v , the transformation function $\Lambda^{v\mu}$ gets as inputs the input $\vec{x}_t$ and cognitive system of frame v , S^v . $\Lambda^{v\mu}$ checks two conditions, whether $\vec{x}_t$ is a state of the state space of C^v (the functional consciousness subsystem of frame v), and whether $\vec{x}_t$ is in $\vec{r}_0^v$, the spatial position of cognitive system S^v (as seen from its own cognitive frame of reference, v). The spatial position of $\vec{x}_t$ as seen from cognitive frame of reference v , is $\vec{r}_{x_t}^v$. Spatial positions $\vec{r}_0^v$, $\vec{r}_{x_t}^v$ are part of the physical properties of the cognitive system of frame v and the input (respectively). The transformation function checks the first condition by plugging $\vec{x}_t$ into m_C^v (that checks whether the input is a state of the state space of C in the cognitive frame of reference v) and checks the second condition by using a delta function, $\delta^{\vec{r}_0^v \vec{r}_{x_t}^v}$, that equals 1 if $\vec{r}_{x_t}^v = \vec{r}_0^v$ and 0 if $\vec{r}_{x_t}^v \neq \vec{r}_0^v$. If $\vec{x}_t$ is not a state in the space state of C^v , then m_C^v will return the empty set. This means that the input $\vec{x}_t$ cannot be measured directly by the functional consciousness subsystem of frame v . The same situation occurs if the delta function is equal to 0, it means that $\vec{x}_t$ is not in the cognitive system's spatial position and thus is not available for C^v to use it and measure it directly. As a result, $\vec{x}_t$ is sent as an input to cognitive system S^v for an indirect measure. There, the sensation subsystem of frame v gets $\vec{x}_t$ as input. From

there, the process will continue along the hierarchy of the cognitive system to the perception subsystem and the functional consciousness subsystem of frame v (see eq. 2, 5). If $\vec{x}_t$ is a state in the state space of C^v , then m_C^v will not return the empty set. This means that the input can be measured directly by the functional consciousness subsystem of frame v . In addition, the second condition is being checked and if the delta function is equal to 1, it means that the input is in the spatial position of the cognitive system and thus available for C^v . As a result, if the two conditions are fulfilled, $\vec{x}_t$ is sent as an input directly to C^v , the functional consciousness subsystem of cognitive system S of frame v , to measure the input directly as a quale. In sum, cognitive system S of frame v is in the third-person perspective if either $\delta_0^{v\vec{r}_{x_t}} = 0$ or $m_C^v(\vec{x}_t) = \emptyset$ because it uses the sensation subsystem (condition for third-person frame of reference is fulfilled, eq. 43) and cognitive system S of frame v is in the first-person perspective only if both $\delta_0^{v\vec{r}_{x_t}} = 1$ and $m_C^v(\vec{x}_t) \neq \emptyset$, because then it uses $C^v(\vec{x}_t)$ which measures directly the phenomenal judgment (i.e., a quale. Condition for first-person frame of reference is fulfilled, eq. 42).

Let's check if Equation 46 gives the correct measurements for each case of the transformation from frame μ to v while preserving the form of the transformation equation (Equation 44). We mentioned three cases: the cognitive frames can measure a third input, (e.g., apple), they can measure frame μ , or they can measure frame v . In the first case, we start from a quale of frame μ about a third input like an apple, $\vec{Q}_{x_t}^\mu$. The transformation function checks if the input $\vec{x}_t$ is a state from the state space of C^v or is it inside the cognitive system S^v . Because this is an input from outside of the functional consciousness subsystem and the cognitive system, $\vec{x}_t$ (the apple) is not a state of the subsystem and not in the spatial position of the cognitive system, and hence the transformation function sends $\vec{x}_t$ to S^v , the cognitive system in frame v . As a result, S^v will use its sensation, perception, and functional consciousness subsystems to create its own quale of the apple, $\vec{Q}_{x_t}^v$. The transformation function succeeds in giving the correct output for the first case. In the second case, frame μ measures its own quale $\vec{x}_t = \vec{Q}_{x_t}^\mu$ and thus in the transformation equation (eq. 44), $\vec{Q}_{x_t}^\mu = \vec{Q}_{Q_{x_t}^\mu}^\mu$. The transformation function checks if the input $\vec{x}_t$ is a state from the state space of C^v and if it is part of the cognitive system S^v . Because this is again an input from outside of the cognitive system and the functional consciousness subsystem of frame v (it's a quale from frame μ), the transformation function sends the input to S^v . As a result, frame v will use its sensation, perception, and functional consciousness subsystems to create a quale of the physical properties of the quale from frame μ , and thus in the transformation equation (eq. 44) $\vec{Q}_{x_t}^v = \vec{Q}_{\sum \alpha_m \hat{e}_m}^v$. The transformation function succeeds in giving the correct output for the second case as well. In the last case, frame μ measures a quale $\vec{x}_t = \vec{Q}_{x_t}^v$ of frame v . As a result, frame μ will measure the physical properties of the quale from frame v , and thus in the transformation equation

(eq. 44), $\bar{Q}_{x_t}^\mu = \bar{Q}_{\sum \alpha_m \hat{e}_m}^\mu$. The transformation function checks if the input $\bar{x}_t$ is a state from the state space of C^ν and if it's part of the cognitive system S^ν . This time, the answer is positive for both conditions. Because this is an input from within the functional consciousness subsystem of frame ν , the matching function of the subsystem m_C^ν will return something other than the empty set and the delta function will be equal to 1.

Consequently, the transformation function sends $\bar{x}_t$ directly to the functional consciousness subsystem of

frame ν [where, $C^\nu(\bar{x}_t) = C^\nu(\sum \alpha_m \hat{e}_m) = C^\nu(\bar{c}_{\bar{x}_t}) = \bar{c}_{\bar{x}_t}^\nu$, see Equation 45]. As a result, frame ν will measure its own quale, and thus in the transformation equation (eq.

44) $\bar{Q}_{x_t}^\nu = \bar{Q}_{Q_{\bar{x}_t}}^\nu$. The transformation function also succeeds in giving the correct output for the last case. Notice that the transformation Equation 44 maintains the same form regardless of the specific measurement case or which cognitive frames of reference are involved. Hence, Equation 44 satisfies the relativity principle as desired.

All the preceding scenarios can give the impression that the two conditions are redundant. They always seem to agree with each other and so, one condition is enough to get the correct transformation. But now let's choose a special case where $\nu = \mu$.

Hence, the transformation function will be $\Lambda^{\nu\nu}(\bar{x}_t, S^\nu) = \Lambda^{\mu\mu}(\bar{x}_t, S^\mu)$. There are two scenarios for this case. One scenario is the identity transformation in which the transformation leaves us in the same cognitive system that we started from (and thus in the same cognitive frame of reference). According to Equation 46, $\Lambda^{\mu\mu}$ gives the correct answer in this scenario, because we stay in the same cognitive system all the time (same frame and same spatial location). In other words, the transformation function doesn't change anything, $\Lambda^{\mu\mu} = 1 \Leftrightarrow \Lambda^{\nu\mu} = \delta^{\nu\mu}$. We can call this the full symmetry scenario, because the two cognitive frames are completely symmetric and we cannot distinguish between them (in other words, we started from a cognitive system and come back after the transformation to the same system). The transformation Equation 44 will be in this scenario,

$$\bar{Q}_{x_t}^\nu = \delta^{\nu\mu}(\bar{Q}_{x_t}^\mu). \quad (47)$$

The second scenario, however, is more complicated. Here we have two cognitive systems but they are in the same cognitive frame of reference. They have the same dynamics and thus the same representations and the same qualia, but they remain two discrete systems with different spatial positions. As a result, for the same input both will have the same answer to every question they might be asked. This is the scenario that we started from when we developed the equivalence principle between ALICE and Alice. It's very special scenario where we have two cognitive systems in different spatial positions but in the same cognitive frame. Notice, however, that even though they are in the same cognitive frame, because they have different spatial positions, they cannot measure directly the representations of the other system and thus they must use their sensation subsystem and measure indirectly the quale of the other system (similar to Equation 41 where the same cognitive system will still measure the physical properties

of its own quale due to the sensation subsystem). In other words, although they are in the same cognitive frame of reference, Alice will measure ALICE from a third-person perspective and vice versa. Will Equation 46 give the correct transformation from frame μ (ALICE) to v (Alice) also in this scenario?

For the first case, both Alice and ALICE measure a third input like an apple. In this case, when we transfer from ALICE to Alice, we will get an answer just like in the previous, full symmetric scenario. Because the physical properties of the apple is from the outside of the cognitive system S^v , the transformation function will use the sensation, perception, and functional consciousness subsystems of the system in frame v to create the

appropriate quale, $\vec{Q}_{x_t}^v$. But because both systems are in the same cognitive frame, their qualia are the same and the quale of the input will remain the same. As a result, the transformation gives us the correct answer for this case,

$$\vec{Q}_{x_t^*}^v = \delta^{v\mu} (\vec{Q}_{x_t^*}^\mu), \quad (48)$$

where $\vec{x}_t^*$ is input from outside of both cognitive frames. ALICE and Alice will measure the same quale about an input that is outside of both their cognitive frames.

If ALICE (μ) measures its own quale, $\vec{Q}_{x_t}^\mu = \vec{Q}_{\vec{x}_t}^\mu$, the transformation function checks if the input $\vec{x}_t = \vec{Q}_{\vec{x}_t}^\mu$ is a state from the state space of C^v and if it's in the position of the cognitive system S^v . In this case, because $\mu = v$, $\vec{x}_t$ is a state of the functional consciousness subsystem of frame v as well (the state $\vec{x}_t = \sum_{m=1}^j \alpha_m \hat{e}_m$ is the same state in both frames). But the other condition fails because the systems are not in the same spatial position, and $\vec{x}_t$ comes from the position of the other cognitive system, S^μ . As a result, $m_C^v(\vec{x}_t) \neq \emptyset$, but $\delta^{\vec{r}_0^v \vec{r}_{x_t}^\mu} = 0$. Because only one condition is fulfilled for the first-person perspective and not both, the transformation function will send $\vec{x}_t$ to S^v for an indirect measure. From the sensation subsystem in frame v until eventually creating a quale of the physical properties of the quale from frame

μ , $\vec{Q}_{x_t}^v = \vec{Q}_{\sum \alpha_m \hat{e}_m}^v$. The transformation function succeeds in giving the correct answer also for this case. If ALICE (μ) tries to measure the quale of Alice (v) the result will be a

quale of the physical properties of Alice's quale, $\vec{Q}_{x_t}^\mu = \vec{Q}_{\sum \alpha_m \hat{e}_m}^\mu$. The transformation function checks if the physical properties of the input $\vec{x}_t = \sum_{m=1}^j \alpha_m \hat{e}_m$ is a state from the state space of C^v and if it's in the position of the cognitive system S^v of frame v (Alice). This time, the answer is positive for both conditions. Because the input is also a state from the state space of the functional consciousness subsystem of frame v and it is in the position of S^v , the matching function of the subsystem m_C^v will return something other than the empty set and the delta function will be equal to 1.

Consequently, the transformation function sends $\vec{x}_t$ directly to the functional consciousness subsystem of the cognitive system in frame v . As a result, frame v will

measure its own quale, $\bar{Q}_{x_t}^\nu = \bar{Q}_{Q_{x_t}}^\nu$. The transformation function also succeeds in giving the correct output for the last case. Now we see that the last two cases in the scenario of ALICE and Alice are different than the full symmetry scenario. The different spatial positions of the cognitive systems break the symmetry of the identity transformation. Although the systems are in the same cognitive frame, the different positions cause the transformation to be different than $\delta^{\nu\mu}$. The different spatial positions break the symmetry because they cause the cognitive systems to use their sensation subsystem to measure the quale of the other system. We can call this scenario the broken symmetry scenario. As a result, we can write the transformation Equation 44 for this scenario as:

$$\bar{Q}_{x_t}^{\nu_{r_0}} = \delta^{\nu\mu} \Lambda^{\nu_{r_0} \mu_{r'_0}} \left(\bar{Q}_{x_t}^{\mu_{r'_0}} \right), \quad (49)$$

Which gives us -

$$\bar{Q}_{x_t}^{\nu_{r_0}} = \Lambda^{\nu_{r_0} \nu_{r'_0}} \left(\bar{Q}_{x_t}^{\nu_{r'_0}} \right). \quad (50)$$

Here $\bar{x}_t^{**}$ is an input of a state from the state space of C either from frame ν or μ in the broken symmetry case. The delta function is presented because both cognitive systems are in the same frame, but $\Lambda^{\nu_{r_0} \nu_{r'_0}}$ is still different than the identity transformation because the symmetry of the cognitive systems was broken by their different spatial positions, r_0, r'_0 (for another example of the importance of the spatial positions condition, see [Supplementary Note 3](#)). Because of the broken symmetry scenario, we define the first-person frame of reference slightly differently than the definition of a cognitive frame of reference. The first-person frame of reference is defined not only by the dynamics of the cognitive system but also by its position. As a result, Both Alice and ALICE will have the same answers to the question of what is it like (because they are in the same cognitive frame of reference), but because they are not in the same spatial position, they are not in the same first-person frame of reference. When one is measuring the dynamics of the other, all they see is the physical patterns of the other cognitive system. In other words, when one cognitive frame is in another spatial position, it has to use its sensation subsystem and thus satisfy the condition for the third-person frame of reference (eq. 43). Only if the two identical cognitive systems can be in the same position can they share the same dynamics and use their functional consciousness subsystem directly without the need to use the sensation subsystem. Only then will they satisfy the condition of the first-person frame of reference (eq. 42).

Now we also see that the general transformation, $\tilde{\Lambda}^{\nu\mu}$, that we used for the equivalence principle in order to prove eq. 35 is a special case of the transformation function we developed here, $\Lambda^{\nu\mu}$. In this case, the two cognitive frames of reference are the same but have different spatial positions (e.g., Alice and ALICE), and their input is constrained to be a question (the same question for both cognitive frames of reference). Because

both frames are the same and because their input is the same question, we always get a scenario like eq. 48 and thus, $\tilde{\Lambda}^{v\mu} = \delta^{v\mu}$.

The conclusion is that the transformation function $\Lambda^{v\mu}$ gives the correct answers to all cases and scenarios while preserving the form of the transformation equation (equation 44) as desired. Equation 46 for $\Lambda^{v\mu}$ can be approximated by a simpler form:

$$\Lambda^{v\mu}(\vec{x}_t, S^v) = \left(1 - \delta^{\vec{r}_0^v \vec{r}_{x_t}^v}\right) S^v(\vec{x}_t) + \delta^{\vec{r}_0^v \vec{r}_{x_t}^v} C^v(\vec{x}_t) \quad (51)$$

The first term is the condition to be in a third person perspective because of the use of the sensation subsystem. Only when $\delta^{\vec{r}_0^v \vec{r}_{x_t}^v} = 0$ will this term be equal to 1. The second term is the condition of being in the first-person perspective (because of the use of the functional consciousness subsystem). Only when $\delta^{\vec{r}_0^v \vec{r}_{x_t}^v} = 1$ will this term be equal to 1. This approximation checks only if the spatial position of the input is the same as the spatial position of the cognitive system and it gives the correct answers to all scenarios. It transforms from the third-person frame of reference to the first-person frame of reference and vice versa. But it is only an approximation because it does not take into account the matching function $m_C(\vec{x}_t)$ and so in the case of two cognitive systems from the same cognitive frames of reference (like Alice and ALICE) the approximation will neglect this information and the result will be two different cognitive frames v, μ instead of just one.

In [Supplementary Note 4](#), we show the inverse transformation function, from frame of reference v to frame μ and that the transformation function and the inverse transformation function cancel each other out (giving the identity transformation), as expected.

Discussion

Consciousness as a Relativistic Phenomenon

As we saw, the assumption that the relativity principle also includes measurements of cognitive systems forces us to treat consciousness as a relativistic phenomenon. If the relativity principle is true, then zombies are not nomologically possible (see Equations 18, 32). In other words, they are not consistent with the extant laws of nature (note that this is a different claim from logical possibility). Recall that both naturalistic dualism and illusionism can be understood to be opposing responses to the same basic paradox of phenomenal judgment: phenomenal judgments seem to not require phenomenal consciousness. This opens the possibility for zombies. [Chalmers \(2017, 1996\)](#) takes the threat of zombies to force us to accept that phenomenal properties are an additional fundamental component of reality. Illusionists instead take *us* to be the zombies ([Frankish, 2017](#); [Graziano et al., 2020](#)). But we have demonstrated that zombies are not nomologically possible in a relativistic universe. The illusionist can, of course, point out that we have assumed that Alice is phenomenally conscious, which they do not grant.

While this is an assumption (and not a controversial one for any but the illusionist), what we point out is that the illusionist contention that there are zombies, and that we are the zombies, is premised on the paradox of phenomenal judgment. But we have eliminated that paradox by eliminating the nomological possibility of zombies. Therefore, we have eliminated the illusionist motivation to deny that Alice is phenomenally conscious. While we can't prove that Alice is indeed phenomenally conscious, the illusionist no longer has a reason (viz., the paradox of phenomenal judgment) to doubt it.

As a result of the relativity principle, there is a formal equivalence between functional consciousness (creating phenomenal judgments) and phenomenal consciousness (qualia and eidetic structures; see Equation 35). Ultimately, we are returning to [Nagel's \(1974\)](#) definition of consciousness: that the creature is conscious if there is something that is like to be this creature. If an observer has phenomenal judgments, then the observer measures that there is something that it is like to be that observer, and this observer is conscious. The essence of the relativistic principle is that there is nothing over and above the observer. Hence, in contrast to illusionism and naturalistic dualism, there cannot be any third-person or God's-eye perspective telling us that some observer is delusional and doesn't *really* have consciousness.

Moreover, phenomenal properties are no longer absolute determinations, but depend on the observer's cognitive frame of reference, just as in the case of constant velocity and the question of who's stationary and who's moving. In a non-relativistic universe, an object O evinces either property P or $\neg P$. Properties are absolute determinations. But in a relativistic universe, an object O evinces either property P or $\neg P$ *with respect to some observer X* . In the first-person cognitive frame of reference, quale Q (for example) is observable, while any third-person cognitive frame of reference observes $\neg Q$, i.e., that there is no quale.

Cognitive frames of reference are defined by the dynamics of the cognitive systems involved (according to the three-tier information processing model, eqs. 2–10). The first-person frame of reference takes into account also the position of the dynamics (eqs. 49–50). The first-person frame of reference of Alice, for example, is the position from within Alice's cognitive system, which is a physical position in space and time where the dynamics of the system take place. This is the only situation that satisfies the condition for the first-person frame of reference (eq. 42). For that reason, the transformation function $\Lambda^{\nu\mu}$ (that transforms between cognitive frames of reference and between first and third frames of reference) takes into account not only whether the input is recognizable by the cognitive system as one of its own representations, but also if the input is in the position of the cognitive system (eq. 46). The third person frames of reference are the positions of other cognitive systems, like Bob, that measure Alice's cognitive system. While Alice may observe herself feeling happiness as she's looking at a rose, Bob will only measure patterns of neural activity. Recall the case of constant velocity, wherein Alice claims to be at rest while Bob is moving, while Bob claims that Alice is the one moving and that he is stationary. From Alice's perspective, she has qualia and Bob only has patterns of neural activation, while from Bob's perspective Alice just has patterns of neural activation while he has qualia. In other words, just as in the case of constant velocity, Bob and Alice both measure that they are the stationary ones,

and hence will not agree on who is stationary and who is moving, there is likewise a relativistic equivalence between all cognitive systems that have phenomenal consciousness or functional consciousness. Both Alice and Bob will measure that they are the ones who have phenomenal consciousness while the other has only neural patterns and hence will not agree on who has phenomenal consciousness.

Alice and Bob will continue to argue over who is right. As a result, Alice might be an illusionist and claim not only that Bob is delusional about having qualia, but moreover that qualia don't exist at all. And, as a response, Bob might claim that qualia are uniquely private and non-physical phenomena. But, just as with the constant velocity case, their conclusions are wrong because they don't grasp the relativistic principle. The illusionist mistake is to claim that the third-person perspective is the only legitimate perspective. When Alice observes Bob's neural firing, he supposedly should infer that all his "qualia" really are just neural firing. Yet the first-person perspective is also a physical arrangement of a cognitive system. It is the cognitive system from its own observer perspective. Both frames of reference are equivalent, there is no observer position that is privileged, and it is impossible to privilege one observer cognitive position over the other. Alice's perspective is not some absolute, "correct" perspective that sets the standard for measurement. Bob doesn't have phenomenal consciousness relative to Alice's perspective but *does* have phenomenal consciousness relative to Bob's perspective. In other words, Bob evinces $\neg Q$ with respect to Alice, and Bob evinces Q with respect to Bob. This is not a logical contradiction, but a statement of the relativity of properties.

On the other hand, qualia and eidetic structures are not private and hence phenomenal consciousness is neither some mysterious force beyond the realm of science nor an irreducible element of reality. Rather, they *appear* to be private because in order to measure them, one needs to be in the appropriate frame of reference, viz., that of the cognitive system in question (see Equations 42, 45). It is ultimately a question of causal power. Only from this frame of reference is there causal power for the representations in the dynamics of the system. Only from the frame of reference of the cognitive system are these neural patterns recognized as representations (eq. 45). These representations are the input and output of the cognitive system (eq. 2-10); they are the ones that cause the dynamics in this cognitive frame of reference. From outside that observer reference frame, as in the position the neuroscientist takes as a third-person observer, the same exact phenomena appear as neural computations. This is because the third-person perspective is constitutively outside of the dynamics of representations of the cognitive system in question, and hence these representations do not have any causal power over the neuroscientist (Equations 36–38). According to the equivalence principle, when Alice observes herself to be happy, it is because her cognitive system can recognize and use the appropriate phenomenal judgments, and these judgments are equivalent to phenomenal consciousness (Equation 35), because there is always an equivalent phenomenal judgment representation for every phenomenal property, and we cannot distinguish between them. These representations and their relations cause the cognitive dynamical system to react with new representations and instantiate new relationships between them. As a result, we get a dynamical system that uses *very specific* representations as variables and as outputs. Any other cognitive system, like

that of Bob, uses *different* representations and thus cannot use the representations of Alice directly in its own dynamics (Equations 36–38). The only possibility left for Bob is to process Alice’s representations through his sensory system and build his own representations. Consequently, we get a sharp difference between the self-measurements of the cognitive system (first-person perspective) and measurements from the outside the cognitive system (third-person perspective), which are mediated solely through the three-tier hierarchy—from the sensation subsystem to the perception subsystem and on to the functional consciousness subsystem (Equations 39, 40). Ultimately, the reason for the sharp difference between first-person and third-person frames of reference is because the sensation subsystem can observe only the physical properties (y_i) of inputs from the outside of the cognitive system, while the functional consciousness subsystem can only observe representations from within the cognitive system (Equations 41–43). Even if the two cognitive systems at hand are in the same cognitive frame of reference (which mean that they have the same dynamics), because they have different spatial positions where their dynamics take place, they will not be in the same first-person frame of reference. They will still need to use their sensation subsystem to measure the other system (hence satisfying the condition of the third-person frame of reference, eq. 43).

Phenomenal properties are not truly private. They seem private because it is non-trivial to do the transformation to the appropriate cognitive frame of reference. Nevertheless, this kind of transformation is nomologically possible. We showed that, as a result of the relativity principle, there is a transformation between the qualia of all cognitive frames of reference (Equations 44, 46, 51). Using this transformation, we obtained an equation that agrees with the relativistic principle (Equation 44). It describes the laws of physics with the same form in all admissible cognitive frames of reference. In other words, this form will stay the same regardless of which cognitive frame we choose (it doesn’t matter for the equation what the specific μ and ν cognitive frames are). The equation enables us to move from the phenomenal consciousness of one cognitive frame of reference to the phenomenal consciousness of another frame and from first-person frame of reference to third person frame of reference (and vice versa). According to the transformation, each observer can change cognitive frames of reference to any other frame by changing the dynamics of its cognitive system. The transformation equation is built in such a way that ensures that after we applied the transformation and moved from one frame to the other, we will get the correct values of the measurements in the new cognitive frame of reference. For example, the equation ensures that the relations between first- and third-person frames of reference are always satisfied. The observer will always measure qualia and eidetic structures from within its own frame, and brain patterns (or any other physical patterns that govern the cognitive system) for all other frames. The only way to enter the frame of reference of the cognitive system is to have the dynamics of representations of that system, because they only have the right kind of causal power from within that cognitive system (they are the right “fuel” of the cognitive system – the inputs and outputs of the system). Hence, third-person studies will get different measurements than those taken from within the system itself, unless a proper transformation were to change the frame of reference of the third-person cognitive system to have the exact dynamics of the cognitive system in question (see Equations 36–38 for third-person frame, in contrast to Equations 42, 45 for the first-person frame).

After such a transformation, there would be an equivalence between the two systems and they would observe the same conscious experience. Obviously, no such transformation is currently technologically feasible, but it is nomologically possible. Because of this transformation that preserves the form of the measurements, phenomenal structures (first-person perspective) and phenomenal judgement-representations (third-person perspective) are equivalent and isomorphic to each other. They are not equal, but they have the same preserved mathematical form such that we can map between them. All that separates them are different kinds of measurements (causing different kinds of properties).

The Hard Problem Dissolves

The relativistic theory of consciousness dissolves the hard problem. If there is no irreducibly private property of phenomenal features, there is no need for adding new, exotic elements to reality, nor is there a need to explain phenomenal features away as illusory. There is also no explanatory gap, because there is no need to explain how physical patterns in the brain create private, irreducible properties. Because consciousness is a relativistic phenomenon, physical patterns (e.g., of neural representations) and phenomenal properties (e.g., qualia) are two sides of the same coin. Both are valid physical descriptions of the same phenomenon from different frames of reference. Although phenomenal properties supervene on physical patterns (the representations that we called 'phenomenal judgements'), they are not created by the physical patterns. Instead, phenomenal properties are the result of a special measurement of these physical representations by the observer. For this observer, these representations are the ones that cause its own dynamics. Notice that according to the relativistic theory of consciousness, the opposite is also true and physical patterns (phenomenal judgements) supervene on phenomenal properties. They are not created by phenomenal properties; instead, physical representations are the result of a measurement of these phenomenal properties by a different observer. For this cognitive frame of reference, these phenomenal properties have no causal power over its dynamics. Phenomenal and physical aspects supervene on each other. There is a subtle identity between them (i.e., they are equivalent). They are just different perspectives of the same phenomenon from different cognitive frames of reference. As a result, the relativistic theory of consciousness is a physicalist theory, but not a reductive theory, because there is no reduction of phenomenal properties to brain patterns. For phenomenal properties we need a cognitive system with the right kind of representations and the right kind of measurement.

In the introduction, we raised a concern about whether any physicalist theory can solve the hard problem of consciousness. How can there be an identity between public properties like structure and dynamics and between private qualia? The answer is that we didn't include relativity as a part of physicalism. The relativistic theory of consciousness shows that phenomenal states are not private and gives an explanation of why they are different from other physical states. Different kinds of measurements give rise to different properties. While our sensation subsystem can only directly measure a physical substrate via measurement devices like the retina, the perceptual and cognitive subsystems directly measure only the roles and relations of

representations within their subsystems and cannot directly measure the physical substrate serving as the referent of their representations (that's why in the equations that describe the dynamics of the perceptual and cognitive subsystems, the inputs are representations and not the physical substrate of the representations, while in the sensation system the inputs are the physical substrates themselves like light or sound waves). From these different kinds of measurements different kinds of properties arise.

[Chalmers \(1996, 2017\)](#) parses the zombie argument in terms of the logical possibility of zombies. He would not deny that zombies are inconsistent with the extant laws of nature, because he speaks only about logical possibility in general. However, the relativistic theory of consciousness carries ontological constraints and conditions about the existence of consciousness in every possible world. For example, the difference between first- and third-person frames of reference arise because of the difference between the kind of measurements that can be taken within each frame of reference. There could be a possible world, not governed by the known laws of nature, wherein the relativistic principle is false. But then, there would be no different frames of reference, and the different kinds of measurements would yield the same result. In other words, there would be no first-person and third-person frames of reference. In such a possible world, we would be left with either the illusionist view, with no consciousness at all (i.e., everybody is a zombie with just third-person perspective), or we would be left with a Berkeleian idealist world where there are only phenomenal properties (with shared first-person perspective for everything). Consequently, according to our relativistic theory of consciousness, there cannot be a world that has both phenomenal consciousness *and* zombies. This conclusion undermines the zombie argument. The main goal of the zombie argument was to establish that phenomenal properties cannot be explained reductively in terms of physics. But the relativistic theory does just that, demonstrating that phenomenal consciousness and physics can be reconciled. There might be a logical possibility of zombies, but in such possible worlds, the physical mechanism that allows the transformation to phenomenal properties (i.e., the relativistic principle) must be absent and hence consciousness must be absent all together as well (because all there is left in such a world is the physical substance that constitutes the world). In sum, in non-relativistic worlds without different frames of reference, we could have *either* illusionism (only physical properties) *or* Berkeleian idealism (only phenomenal properties). But in a relativistic world where different frames of reference measure different properties, *both* physical *and* phenomenal properties are possible for the same entity.

Phenomenal consciousness is not private—all we need to do to measure phenomenal properties is to change our perspective by moving to the appropriate frame of reference, i.e., the frame wherein the representations have causal power (i.e., the frame that measures the representations directly). As a result, this frame doesn't measure them merely as physical patterns, but as what they represent and stand for. For example, the representation of an apple has some causal power in the system, according to what has learned about apples. The representation will trigger, for example, memories, emotions, and motor processes. For the cognitive system, this is what it means to be an apple. Thus, the physical properties of the representation are not being directly measured, but rather only its role and relations with other representations in the system. Consequently,

this representation will be measured by the system as an apple and not as a representation of an apple. That is to say, because there is nothing above and beyond the observer, and because the cognitive system (the observer) measures this representation according to what it does (the representation describes all the functions, properties and relationships of an apple in the system), this representation will be measured by the system as an apple. (That's why phenomenal consciousness is characterized by transparency, i.e., we don't perceive representations but directly perceive things). This is a direct measure of the representation itself. In the case of phenomenal judgments representations, these new properties are being measured as phenomenal properties. Specifically, when there is a cognitive frame of reference that creates functional consciousness (phenomenal judgments), it is sensible to assume that such a system will have a special subsystem specializing in such complex representations (the functional consciousness subsystem, C). It uses the complex representations of functional consciousness for its dynamics (i.e., $\bar{c}_{x_i}$ – phenomenal judgment representation, see eq. 5). As a result, these representations have causal power in this frame of reference. The relativity principle ensures us that because no measurement can distinguish between phenomenal properties and their corresponding phenomenal judgments, this cognitive frame will measure these phenomenal judgments as qualia and eidetic structures (Equation 35). In sum, according to the relativistic theory of consciousness, for phenomenal structures we need a cognitive system with the right kind of representations and the right kind of measurement. Phenomenal structures are the direct measurements of the complex representations that we called phenomenal judgments.

The relativistic theory of consciousness suggests a solution for the hard problem based in relativistic physics. There are still several open questions that need to be addressed in the future. For example, what are the necessary and sufficient conditions for a cognitive frame of reference to have phenomenal consciousness? And what empirical predictions does this theory yield? On the philosophical side, there are also questions about the affinities and differences between our view and dual aspect monisms such as reflexive monism ([Velmans, 2009](#)). These are issues that we will address looking forward.

The Formal Equivalence of Phenomenal and Neurocomputational Structures

Since consciousness is relativistic, there is a formal equivalence between first-person phenomenological structures and third-person neurocomputational structures. Philosophers have asked and puzzled over the question, "What is it like to be a bat?" ([Nagel, 1974](#)). Nagel's point is that physicalism is not sufficient for explaining consciousness. Even if we describe the neural processes involved in bat sonar, we will never get an intuitive or imagistic sense of what it is like from the first-person perspective. Yet years of philosophical arguments about the impossibility of a complete scientific understanding of consciousness has only stymied serious research, relegating consciousness to something that is supposedly not measurable and hence not amenable to scientific treatment ([Holland, 2020](#)).

By the relativistic theory of consciousness, the neurocomputational structures are equivalent to the phenomenal structures of what it is like to be a bat. This equivalence is not a statement of the reductive physicalism that Nagel attacks. We do not maintain that phenomenology can be reduced to neural computation. Rather, we maintain that phenomenology and neural computation are two different ways that the same phenomenon appears based on the cognitive perspective of the observer. That perspective is either from within the cognitive system (the first-person perspective) or outside of the cognitive system (the third-person perspective). This equivalence allows us to use neurocomputational structure to derive phenomenal structure. In other words, phenomenal consciousness can be investigated by studying neurocomputational dynamics. Certainly, the only way that the intuitive, imagistic aspects of “what it is like” could be perceived is by actually taking the cognitive frame of reference of the system in question. But qualia and eidetic structures can be measured indirectly through the neural phenomena they manifest from the third-person cognitive frame of reference. Phenomenal consciousness is not an immaterial and extra-scientific phenomenon, but one that’s amenable to the scientific method (for an example of a similar approach, see [Petitot, 2018](#)).

Not only can neurocomputational structure help us discover phenomenological structure. Furthermore, phenomenology can also help us to discover neurocomputational structures themselves. We are referring to neurophenomenology ([Varela, 1999](#); [Gallagher, 2003](#); [Varela and Thompson, 2003](#); [Petitmengin, 2006](#); [Petitot, 2018](#); [Gallagher and Zahavi, 2020](#)). Rigorous first-person philosophical inquiry into phenomenal structure can help guide cognitive neuroscience research into consciousness by tracing the lineaments of what scientists ought to be looking for. While descriptions of phenomenal consciousness are often thought of as wispy and vague, there are precise ways that phenomenological philosophy (and so-called “cognitive phenomenology”) describes phenomenal consciousness that go beyond naïve, folk conceptions of experience.

For the illusionist, who understands phenomenal consciousness to be a cognitive illusion of phenomenal properties, “cognitive scientists should treat phenomenological reports as fictions” ([Frankish, 2017](#), p. 27). In the relativistic perspective, however, phenomenological reports are not fictions. Of course, there are serious methodological issues that make data collection from the first-person perspective the purview of philosophy rather than science ([Husserl, 1925/2012](#)). But phenomenal reports are reports of what the system is like from within its own cognitive frame of reference. These reports can potentially help guide cognitive neuroscientific investigation. Since consciousness is relativistic, phenomenal structure is simply a different way of perceiving the same phenomenon that the cognitive neuroscientist is examining from a different observer perspective. Rather than denigrating the first-person perspective, we believe the relativistic framework invites us to take phenomenological philosophy seriously. We believe it poses potentially rich possibilities for interdisciplinary work on consciousness. No doubt, such interdisciplinary work would be fraught with difficulties, but the phenomenological philosopher looking at phenomenal consciousness and the cognitive neuroscientist looking at functional consciousness are both looking at the same thing from different observer positions.

Let's get a clearer idea on what philosophers can contribute to the empirical study of consciousness. Typically, first-person experience is spoken of in terms of qualia, which often picks out the content of phenomenal consciousness: what it is like to see red ([Jackson, 1982](#)), or what it is like to be a bat and have sonar ([Nagel, 1974](#)). But the tradition of phenomenological philosophy, initiated by Edmund Husserl and continued by Martin Heidegger and Maurice Merleau-Ponty, has long studied first-person experience in terms of eidetic structures. An eidetic structure refers to the “form” or “essence” (*eidos*) of a given phenomenon in first-person experience, that is, its invariant structure across time. Phenomenologists are interested not in the qualia of greenness, but in the structural invariants that are expressed in all phenomenal consciousness, such as intentionality or the temporal structure of cognition and consciousness. We will close this section by describing the eidetic structure of what phenomenologists call time-consciousness ([Husserl, 1917/1991](#); [Neemeh and Gallagher, 2020](#)). We will hypothesize a way that this eidetic structure may be realized as a neurocomputational structure. This should be understood as an invitation for the kind of interdisciplinary exploration between philosophers and cognitive neuroscientists that we have advocated for.

The subjective experience of time has a formal eidetic structure, what [Husserl \(1917/1991\)](#) calls time-consciousness. While we usually imagine time to be a linear arrow, phenomenologists contend that time-consciousness has a far more complicated, looping structure ([Neemeh and Gallagher, 2020](#)). Take the perception of a melody. How can you perceive a melody if, at any given point in time, you only hear a single note being played? Don't you simply perceive individual notes? But a series of individual notes is not the same as hearing a melody. The eidetic structure of melody perception involves what phenomenologists call protention and retention. First of all, the immediate present is never a simple, non-extended slice of time. The present of experience is a “specious present” ([James, 1890/1983](#)) that is temporally extended. When one note is perceived in the immediate present, it recedes into the immediate past. Yet it is still retained, in an attenuated form, in consciousness (called “retention”). It is retained not as a sensuously fulfilled object, but as a trailing shadow ([Husserl, 1917/1991](#)). Only the immediate now is fully sensuous. But if past notes were not retained in this way in consciousness, we would not have a conscious perception of a melody. We would only hear individual notes. At the same time, melody perception involves an anticipation (or protention) of the imminent note. This is how we can perceive a wrong note: it doesn't match up with our anticipation, based on the previously given aspects of the melody. The same principle holds when we hear a sentence.

In the relativistic view, first-person experience is no longer ineluctably private. It is simply one specific cognitive frame of reference, and when we switch perspectives, by the appropriate transformation (eq. 44, 46, 51), to that of cognitive neuroscience (third-person perspective), we ought to (ideally) detect isomorphic structures that are not qualitative but rather quantitative (assuming our phenomenological analysis is correct, of course).

Given this formal equivalence (Equation 44), there is a neurocomputational structure that is equivalent to the eidetic structure of time-consciousness, or perhaps there are

multiple such neurocomputational structures for time-consciousness in different sensory modalities. Cognitive neuroscientific investigation can use this eidetic structure as a guide to inquiry. For example, [Varela \(1999\)](#) suggests an account based in phase locking, and [Neemeh and Gallagher \(2020\)](#) suggest a Bayesian approach.

Philosophers have long studied phenomenal consciousness and may have practiced insight into what structures to look out for in functional consciousness. This is not to say that philosophy is, or has ever pretended to be, a science. We advocate for an interdisciplinary investigation of consciousness that takes eidetic structures as the seeds of empirical hypotheses into neural function (qualia, on the other hand, would likely not be very productive of hypotheses. After all, how much is there to say about redness?). This can open the cognitive neuroscience of consciousness up to vistas of richer views of consciousness.

Conclusion

In sum, we propose a novel, relativistic theory of consciousness, one that accounts for both the functional and phenomenal features of consciousness, bridging the explanatory gap. Through conceptual arguments and mathematical formalizations, we propose that there is no need to expand the basic inventory of nature (as dualists like [Chalmers, 2017](#) argue), nor is there a need to explain away phenomenal features (as illusionists argue). Phenomenal features are not truly private, since the principle of relativity allows us to perform a transformation from one cognitive frame of reference to another. We provided a mathematic transformation between two idealized cognitive systems taken from different cognitive frames of reference, showing their relativistic equivalence. The privacy of phenomenal features is only an illusion, based on our biological limitations and the technological limitations of current science—basically, we can't yet actually perform such a transformation. But our formalization is a proof of concept, showing that it is theoretically feasible. Since phenomenal features are not private, both the presence of zombies and the paradox of phenomenal judgment fall away. The dualist infers from these that phenomenal consciousness is a non-material extra force or property of nature, while the illusionist infers that phenomenal consciousness is merely an illusion created by phenomenal judgments. But once the privacy issue, zombies, and this paradox are neutralized, there is no longer any strong motivation for the dualist and illusionist positions. Phenomenal consciousness is neither private nor delusional, just relativistic.

Not only does the relativistic theory of consciousness legitimize the study of phenomenal features in science, but it furthermore opens up many new questions and possibilities for research. We noted that philosophers studying phenomenal consciousness could play a legitimate role in the science of consciousness, such as by theoretical contributions to experiments seeking the neural basis of phenomenal and eidetic structures.

Data Availability Statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Author Contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Acknowledgments

We would like to acknowledge the Kavli Institute for Cognitive Neuroscience, which generously hosted and led us to develop the ideas in this article, as well as the suggestions of reviewers. NL would like to thank Ashok Vaish for his continuous and generous support. NL would also like to thank Benjamin Pageler, Eva Jablonka, Simona Ginsburg, Avshalom Elitzur, and David Steinhart.

Supplementary Material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.704270/full#supplementary-material>

Footnotes

1. [^] The terms are nearly synonymous, but 'transcendental' means additionally that one has performed the *epochē*. The *epochē* refers to a theoretical bracketing of the ontological status of phenomena—i.e., bracketing the existence of the world. The goal of the *epochē* is to not confuse external reality with the immanent

features of phenomenal consciousness. Since this is not a concern of ours here, we stick with the term 'eidetic.' See [Husserl \(1925/2012\)](#).

References

Alter, T. (2016). The structure and dynamics argument against materialism. *Noûs* 50, 794–815. doi: 10.1111/nous.12134

[CrossRef Full Text](#) | [Google Scholar](#)

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

[Google Scholar](#)

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends Cogn. Sci.* 15, 567–575. doi: 10.1016/j.tics.2011.11.001

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Brown, R. (2010). Deprioritizing the a priori arguments against physicalism. *J. Conscious. Stud.* 17, 47–69.

[Google Scholar](#)

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247. doi: 10.1017/S0140525X00038188

[CrossRef Full Text](#) | [Google Scholar](#)

Carroll, S. (2021). Consciousness and the laws of physics. *J. Conscious. Stud.* 28, 16–31.

[Google Scholar](#)

Chalmers, D. (2017). "Naturalistic dualism," in *The Blackwell Companion to Consciousness*, eds S. Schneider and M. Velmans (Chichester: John Wiley & Sons), 363–373. doi: 10.1002/9781119132363.ch26

[CrossRef Full Text](#) | [Google Scholar](#)

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.

[Google Scholar](#)

Chalmers, D. J. (1996). *The Conscious Mind: In search of a Fundamental Theory*. Oxford: Oxford university press.

[Google Scholar](#)

Chalmers, D. J. (2003). "Consciousness and its place in nature," in *Blackwell Guide to the Philosophy of Mind*, eds S. Stich and T. Warfield (Oxford: Blackwell), 102–142. doi: 10.1002/9780470998762.ch5

[CrossRef Full Text](#) | [Google Scholar](#)

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York: Penguin.

[Google Scholar](#)

Dennett, D. C. (1991). *Consciousness Explained*. New York: Penguin.

[Google Scholar](#)

Dennett, D. C. (1995). The unimagined preposterousness of zombies. *J. Conscious. Stud.* 2, 322–326.

[Google Scholar](#)

Doerig, A., Schurger, A., Hess, K., and Herzog, M. H. (2019). The unfolding argument: why IIT and other causal structure theories cannot explain consciousness. *Conscious. Cogn.* 72, 49–59. doi: 10.1016/j.concog.2019.04.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Dretske, F. (1978). *The Role of the Percept in Visual Cognition*. Minneapolis: University of Minnesota Press.

[Google Scholar](#)

Dretske, F. (2003). "Sensation and perception (1981)," in *Essays on Nonconceptual Content*, ed. Y. H. Gunther (Cambridge: The MIT Press).

[Google Scholar](#)

Einstein, A. (1905). On the electrodynamics of moving bodies. *Ann. Der Phys.* 17, 891–921. doi: 10.1002/andp.19053221004

[CrossRef Full Text](#) | [Google Scholar](#)

Einstein, A. (1907). *Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen*. Stuttgart: S. Hirzel.

[Google Scholar](#)

Einstein, A. (1911). On the influence of gravitation on the propagation of light (M. D. Godfrey, Trans.). *Ann. Phys.* 35, 898–908.

[Google Scholar](#)

Einstein, A., Lorentz, H. A., Minkowski, H., Weyl, H., and Sommerfeld, A. (1923/1952). *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*. Illinois: Dover. (Original work published 1923).

[Google Scholar](#)

Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. (2007). Masking disrupts reentrant processing in human visual cortex. *J. Cogn. Neurosci.* 19, 1488–1497. doi: 10.1162/jocn.2007.19.9.1488

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Forshaw, J., and Smith, G. (2014). *Dynamics and Relativity*. Hoboken: John Wiley & Sons.

[Google Scholar](#)

Frankish, K. (2007). The anti-zombie argument. *Philos. Q.* 57, 650–666. doi: 10.1111/j.1467-9213.2007.510.x

[CrossRef Full Text](#) | [Google Scholar](#)

Frankish, K. (ed.) (2017). *Illusionism As a Theory of Consciousness*. United Kingdom: Andrews UK.

[Google Scholar](#)

Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., et al. (2016). A LIDA cognitive model tutorial. *Biol. Inspired Cogn. Archit.* 16, 105–130. doi: 10.1016/j.bica.2016.04.003

[CrossRef Full Text](#) | [Google Scholar](#)

Gallagher, S. (2003). Phenomenology and experimental design toward a phenomenologically enlightened experimental science. *J. Conscious. Stud.* 10, 85–99.

[Google Scholar](#)

Gallagher, S., and Zahavi, D. (2020). *The Phenomenological Mind*, 3rd Edn. Oxfordshire: Routledge.

[Google Scholar](#)

Graziano, M. S. (2019). *Rethinking Consciousness: A Scientific Theory of Subjective Experience*. New York: WW Norton & Company.

[Google Scholar](#)

Graziano, M. S., Guterstam, A., Bio, B. J., and Wilterson, A. I. (2020). Toward a standard model of consciousness: reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cogn. Neuropsychol.* 37, 155–172. doi: 10.1080/02643294.2019.1670630

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Holland, O. (2020). Forget the Bat. *J. Artif. Intell. Conscious.* 7, 83–93. doi: 10.1142/S2705078520500058

[CrossRef Full Text](#) | [Google Scholar](#)

Husserl, E. (1917/1991). *On the Phenomenology of the Consciousness of Internal Time*. Netherlands: Kluwer.

[Google Scholar](#)

Husserl, E. (1925/2012). *Phenomenological Psychology: Lectures, Summer Semester, 1925*. Berlin: Springer Science & Business Media.

[Google Scholar](#)

Hutto, D. D. (2007). The narrative practice hypothesis: origins and applications of folk psychology. *R. Inst. Philos. Suppl.* 60, 43–68. doi: 10.1017/cbo9780511627903.004

[CrossRef Full Text](#) | [Google Scholar](#)

Jackson, F. (1982). Epiphenomenal qualia. *Philos. Q.* 32, 127–136. doi: 10.2307/2960077

[CrossRef Full Text](#) | [Google Scholar](#)

James, W. (1890/1983). *The Principles of Psychology*. Cambridge: Harvard University Press.

[Google Scholar](#)

Kanai, R., Chang, A., Yu, Y., Magrans, de Abril, I., Biehl, M., et al. (2019). Information generation as a functional basis of consciousness. *Neurosci. Conscious.* 2019:niz016. doi: 10.1093/nc/niz016

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Kanizsa, G. (1985). Seeing and thinking. *Acta Psychol.* 59, 23–33. doi: 10.1016/0001-6918(85)90040-X

[CrossRef Full Text](#) | [Google Scholar](#)

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pac. Philos. Q.* 64, 354–361. doi: 10.1111/j.1468-0114.1983.tb00207.x

[CrossRef Full Text](#) | [Google Scholar](#)

Mashour, G. A., Roelfsema, P., Changeux, J.-P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron* 105, 776–798. doi: 10.1016/j.neuron.2020.01.026

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Mindt, G. (2017). The problem with the 'information' in integrated information theory. *J. Conscious. Stud.* 24, 130–154.

[Google Scholar](#)

Møller, C. (1952). *The Theory of Relativity*, 2nd Edn. Oxford: Oxford University Press.

[Google Scholar](#)

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914

[CrossRef Full Text](#) | [Google Scholar](#)

Nagel, T. (2012). Conceiving the impossible and the mind-body problem. *Rev. Română Filosofie Anal.* 6, 5–21.

[Google Scholar](#)

Neemeh, Z. A., and Gallagher, S. (2020). “The phenomenology and predictive processing of time in depression,” in *The Philosophy and Science of Predictive Processing*, eds D. Mendonça, M. Curado, and S. Gouveia (London: Bloomsbury), 187–207.

[Google Scholar](#)

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Pageler, B. T. (2011). The Explanatory Value of the Cognitive Unconscious. Ph.D. dissertation.

[Google Scholar](#)

Petitmengin, C. (2006). Describing one's subjective experience in the second person: an interview method for the science of consciousness. *Phenomenol. Cogn. Sci.* 5, 229–269. doi: 10.1007/s11097-006-9022-2

[CrossRef Full Text](#) | [Google Scholar](#)

Petitot, J. (2018). *Elements of Neurogeometry*. Berlin: Springer.

[Google Scholar](#)

Pound, R. V., and Rebka, G. A. Jr. (1960). Apparent weight of photons. *Phys. Rev. Lett.* 4:337. doi: 10.1103/PhysRevLett.4.337

[CrossRef Full Text](#) | [Google Scholar](#)

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge: MIT press.

[Google Scholar](#)

Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. New York, NY: WW Norton & Company.

[Google Scholar](#)

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. doi: 10.2307/25470707

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

van Gulick, R. (2014). "Consciousness," in *Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Berlin: Springer).

[Google Scholar](#)

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Varela, F., and Thompson, E. (2003). “Neural synchrony and the unity of mind: A neurophenomenological perspective,” in *The Unity of Consciousness*, ed. A. Cleeremans (Oxford: Oxford University Press).

[Google Scholar](#)

Varela, F. J. (1999). “The specious present: A neurophenomenology of time consciousness,” in *Naturalizing Phenomenology*, eds J. Petitot, F. J. Varela, B. Pachoud, and J.-M. Roy (California: Stanford University Press).

[Google Scholar](#)

Velmans, M. (2009). *Understanding Consciousness*. Routledge.

[Google Scholar](#)

Keywords: consciousness, phenomenology, qualia, relativity (physics), the hard problem of consciousness, mathematical formulization of consciousness

Citation: Lahav N and Neemeh ZA (2022) A Relativistic Theory of Consciousness. *Front. Psychol.* 12:704270. doi: 10.3389/fpsyg.2021.704270

Received: 02 May 2021; **Accepted:** 01 December 2021;
Published: 12 May 2022.

Edited by:

[Peter beim Graben](#), Humboldt University of Berlin, Germany

The finer scale of consciousness: quantum theory

[Tianwen Li](#)^{1,#}, [Hailiang Tang](#)^{1,#}, [Jianhong Zhu](#)^{1,✉}, [John H Zhang](#)^{2,✉}

- Author information
- Article notes
- Copyright and License information

PMCID: PMC6861790 PMID: [31807566](#)

Abstract

Consciousness is a multidisciplinary problem that has puzzled all human beings since the origin of human life. Being defined in various pointcuts by philosophers, biologists, physicists, and neuroscientists, the definitive explanation of consciousness is still suspending. The nature of consciousness has taken great evolution by centering on the behavioral and neuronal correlates of perception and cognition, for example, the theory of Neural Correlates of Consciousness, the Global Workspace Theory, the Integrated Information Theory. While tremendous progress has been achieved, they are not enough if we are to understand even basic facts—how and where does the consciousness emerge. The Quantum mechanics, a thriving branch of physics, has an inseparable relationship with consciousness (e.g., observer effect) since Planck created this subject and its derived quantum consciousness theory can perfectly fill this gap. In this review, we briefly introduce some consciousness hypotheses derived from quantum mechanics and focus on the framework of orchestrated objective reduction (Orch-OR), including its principal points and practicality.

Keywords: Consciousness, quantum mechanics, uncertainty principle, orchestrated objective reduction theory (Orch-OR theory), microtubule

Introduction

The nature of consciousness, once the exclusive realm of philosophers, has been gradually penetrated by neuroscientists, biologists, and physicists. Consciousness has always been defined as the Hard Problem in these subjects ([1-4](#)). With the emergence of unprecedented devices and the development of multidisciplinary experiments in different research fields, more details of this hard problem have been revealed, especially in quantum mechanics and neuroscientific fields. Since we decided to dive into the ocean of consciousness, many may think that the “water molecule” is our best beginning to disentangle this multidisciplinary conundrum. Thus, various interpretations of consciousness originated at the neuronal and molecular levels. The past 30 years have witnessed the tremendous and remarkable achievements of the neural correlates of consciousness ([5,6](#)). Nevertheless, there are many riddles that these “neuronal and molecular” models cannot explain ([7](#)). As a supplement to the theory of macroscopic physics (classical Newton’s mechanics), quantum mechanics has existed only for approximately one hundred years; however, some quantum mechanics terminologies have been ever prevalent, e.g., quantum entanglement, quantum coherence, wave-function collapse, and the most famous cat worldwide—Schrödinger’s cat. Quantum

computers are considered the brightest new star in the quantum field and increasingly fascinate quantum physicists and information technology specialists. Advances in new materials and cryogenic physics have led to remarkable breakthroughs in quantum computing in recent years. Large commercial quantum computer systems that were invented by IBM ([8-10](#)) or the University of Science and Technology of China ([11-13](#)) could soon be operational within five years. Because quantum mechanics deals with the tiniest constituents of the material world, it seems capable of elucidating numerous unsolved and tough problems. Quantum theory, a branch from the finer scale of consciousness, has been accompanied by numerous controversies since its inception, but abundant proof demonstrated that this theoretical framework is capable of explaining the majority of consciousness problems that traditional neuroscience could not, especially the orchestrated objective reduction (Orch-OR) theory introduced by Penrose and Hameroff.

Origination

Consciousness was brought onto the scene that was ablaze with lights for more than 2,500 years since Socrates postulated that the cerebrum created consciousness, while Aristotle argued that mental qualities belonged to fundamental reality. More recently, Descartes proposed “Cogito ergo sum” in the Late Renaissance ([14](#)), and its definition and biological basis remain impasses such that no ultimate version has been reached, although there are various explanations based on its different aspects. Undoubtedly, any objective definition trying to explain this subjective problem seems feeble, which accounts for its official position—a hard problem. Compared with consciousness, quantum mechanics that was first invented approximately one hundred years ago is much younger. However, the latecomers have the upper hand. The last 100 years have witnessed the rapid development of quantum mechanics, and various significant breakthroughs have been accomplished by quantum physicists, while evolution in consciousness has progressed at a snail’s pace.

Initial exploration

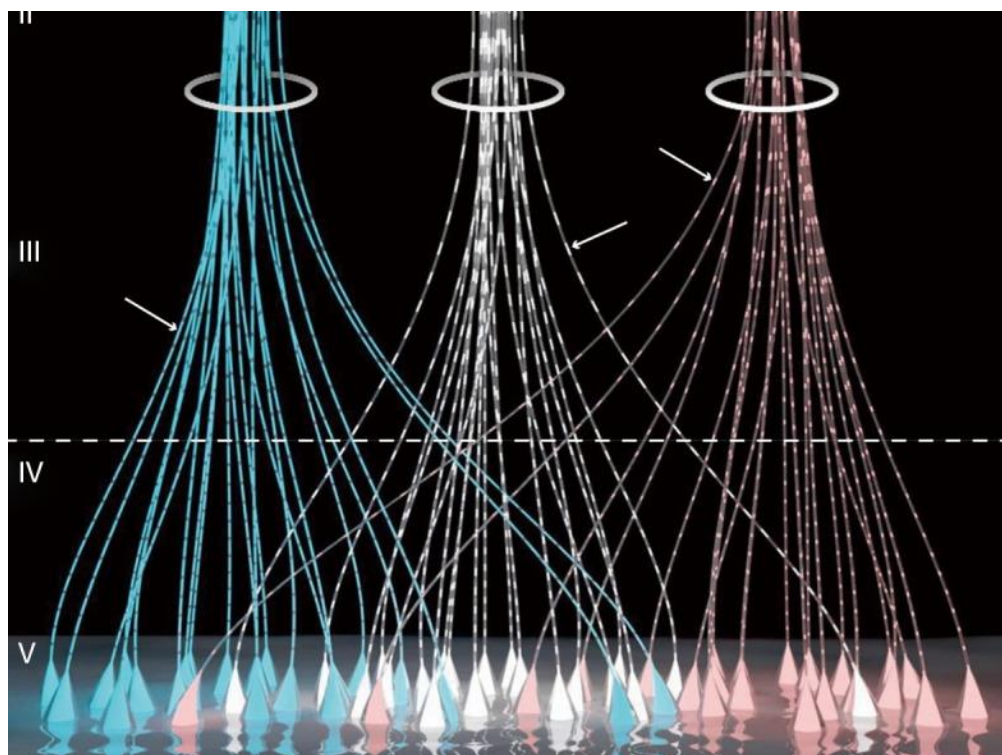
In the 1930s, the Hungarian physicist Eugene Wigner considered it logical that the quantum description of an object is influenced by the mind that comes into our consciousness. In logic, solipsism may be consistent with current quantum mechanics ([15](#)). From the theory of the brain quantum field, Hiroomi Umezawa interpreted memory and other conscious phenomena as the result of the energy exchange of energy particles in the cerebral cortical field ([16-19](#)). Before long, Herbert Fröhlich proposed that the quantum field in the cell membrane condensed into the same state with quantum coherence during charge oscillation, namely, the Bose-Einstein condensate ([20,21](#)). Next, following Fröhlich, Marshall introduced that Bose-Einstein condensate produced from pumped phonons was the candidate for processing consciousness ([22](#)). John Wheeler, a noble American physicist, even suggested that the existence of life, including all life with “observation” ability, may have transformed many of the possible “quantum pasts” into real history. From this perspective, we have been participants since the beginning of the universe. Stunningly but perceptibly, we live in a “participating

universe” (23). In the book of *Quantum Theory* (24), Bohm proposed that implicate order could apply to both the material world and consciousness, which was capable of explaining the relationship between them. In his opinion, mind and matter are projections into the explicate order from the underlying implicate order. By comparing optical holography with visual memory, Gabor attempted to explain the connection between human temporal recall and quantum mechanics (25). Stapp postulated a global collapse, a ‘mind like’ wave-function collapse that exploited certain aspects of the quantum Zeno effect within the synapses (26,27), distinct from the idea of Hameroff and Penrose (described in the “Orch-OR theory” section). Although his theory was based on synapses, a junction structure between neurons within the brain, unfortunately, he did not integrate his exhaustive quantum mechanical theory into an optimal interpretation with neurological or neuronal substrates. Jibu and Yasue believed that the movement of corticon and boson in the dipolar vibrational fields distributed along protein filaments of the cytoskeleton and extracellular matrices and surrounding water fields of the brain was the process of quantum information transmission, which ultimately led to the formation of quantum coherent states in the cortex (28-33). Disagreements and rebuttals were inevitable due to these ideas, and theories were simply overflowing from those physicists’ minds, lacking factual neuroanatomical evidence and even systematic contexture (34-36).

Gradual systematization

Many biological scientists have tried to find a quantum-level substance that can be transmitted between neurons, such as a neurotransmitter, which conveys consciousness-related information (37). Based on “consciousness evolutionism”, which emphasized that it is unlikely that consciousness came to higher creatures as a sudden leap, a finer world had been built. In this microgranular mental world, the “quantum-level substance” is the mental microunits called psychons, wandering around the mammalian neocortex ([Figure 1](#)). Each kind of psychon represents a unitary experience of consciousness-qualia. Like conventional neurotransmitters, psychons are transported along dendrons formed by apical dendrites (also their branches) of pyramidal cells in layers V, III, and II. During the period in which synapse structure was discovered, its foundational hypothesis had no reason to deviate from synaptic micro-property.

Figure 1.



Schematic diagram of dendrons and psychons of layer V pyramidal cells. Each triangle represents a pyramidal cell, and the ascending lines extending from the top of triangles are apical dendrites. Light spots (arrowhead), which are transported along apical dendrites with different colors, represent different kinds of psychons that could give rise to unique experiences. The dendrites cluster automatically while ascending to the superficial layer of cortex to configure dendrons (circle) according to the different psychons they transmit. Additionally, pyramidal cells are divided into different groups based on the distinct conscious experiences that they process ([37](#)).

Following the analogy of probabilistic fields of quantum mechanics and neural events arising from mental activity ([38](#)), as well as the hypothesis of dendrons and synaptic structure mentioned above, a more complete and systematic framework was constructed by Beck ([37](#)). More quantum mechanics mechanisms were also introduced. Exocytosis, as an intriguing discovery, was chosen as the indispensable and pivotal step. In the standby phase of the synaptic bouton, exocytosis was seen as a staged quantal event based on the quantum tunneling effect, resulting in momentarily increasing the probability of exocytosis (actually, it was named the selection of events in quantum mechanics). This selection mechanism augmented the likelihood for exocytosis and led to amplified excitatory postsynaptic potentials. Then, a coherent coupling of boutons in a dendron occurred. Because the dendrons in the cerebral cortex contain countless synaptic connections, there is an enormous amount of quantum transmission in the human brain, which inevitably leads to the randomness and freedom of mental phenomena-free will.

However, after more than twenty years, no subsequent development of this ancestral hypothesis occurred; additionally, the true identity of psychons was not discovered, despite countless works focused on cortical architecture (39). Various limitations persisted in these theories; however, they probably marked the beginning of the convergence of neuroscience and quantum mechanics for further understanding of consciousness.

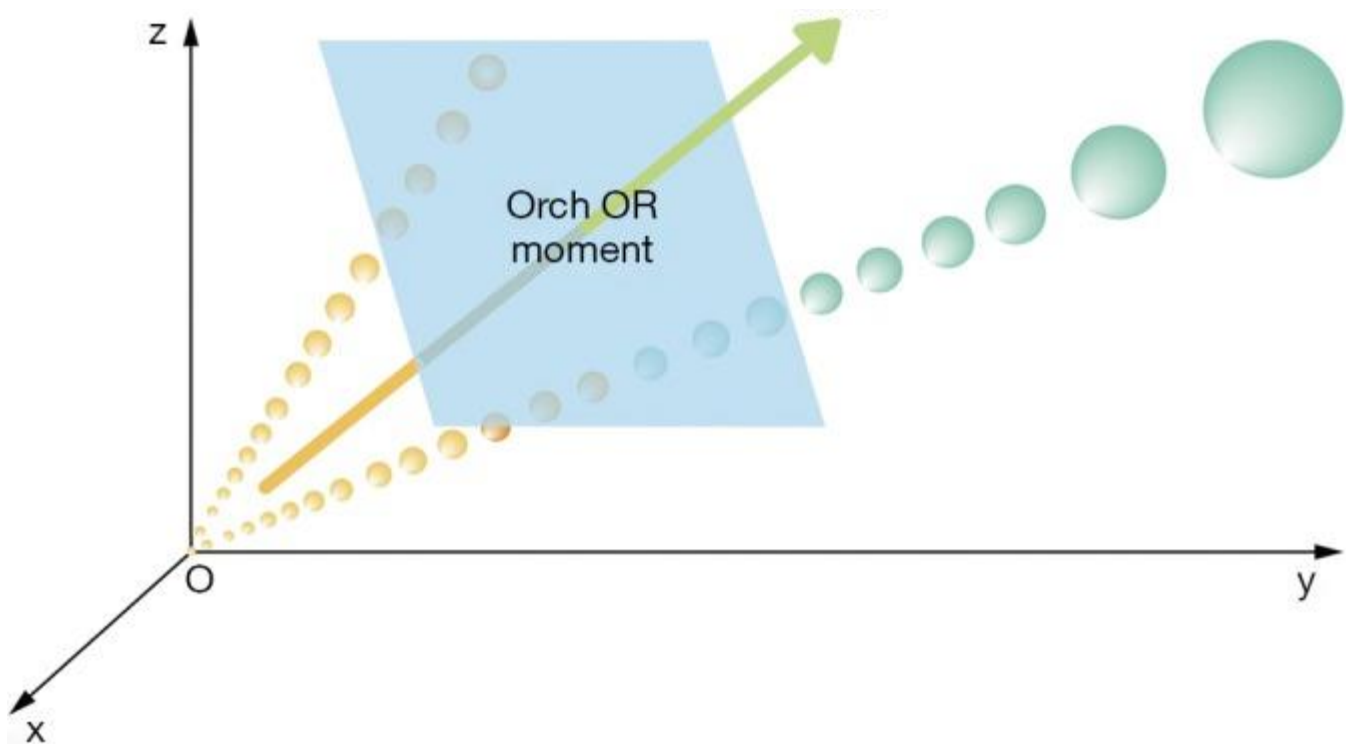
Orch-OR theory

It was widely accepted that most neuronal communication and information transmission initially occurred on receptors and ligands (especially among synapses in the central nervous system) on the cell membrane, followed by second messengers that broadcast or transfer the information to various parts of the interior cell. Almost all basic studies in neurobiology converge on the various receptors, ligands and signaling pathways. However, are we 100% certain about this prerequisite basis of neuroscience? Rather than the conventional receptors and ligands of the membrane, the principal cellular components of the Orch-OR theory are microtubules that are mostly considered pivotal structures for material transportation, cell movement, mitosis and establishment and maintenance of cell form and function.

To date, this theory has remained one of the most acceptable and continuous theories that covers in detail quantum physics, quantum gravity, quantum information theory, molecular biology, neuroscience, cognitive science, philosophy, and anesthesiology. Additionally, this theory was known to neurobiologists who were interested in the “Hard Problem” as well as physicists and philosophers.

Under the background of rapid development of world computer technology, Hameroff likened the flow of information in the brain to computers in which microtubules were to the brain what transistors were to the computer (40-43). Inspired by this fantastic analogy and Gödel's incompleteness theorems, in *The Emperor's New Mind* (44) published in 1989, Roger Penrose first attached the quantum effect in human cognition. For example, he considered whether consciousness can affect quantum mechanics or vice versa and that quantum mechanics itself might be included in consciousness. Penrose suggested that the “objective collapse”, that is, the collapse and superposition of quantum interference, is a real physical process, similar to the bursting of bubbles (44). Furthermore, consciousness was the product of quantum space-time structure (Figure 2), which was inextricably related to the universe, and the theory describing the relationship between consciousness and the universe was the Orch-OR theory (44). These quantum theories facilitated the emergence of later biological hypotheses of consciousness based on quantum mechanics.

Figure 2.

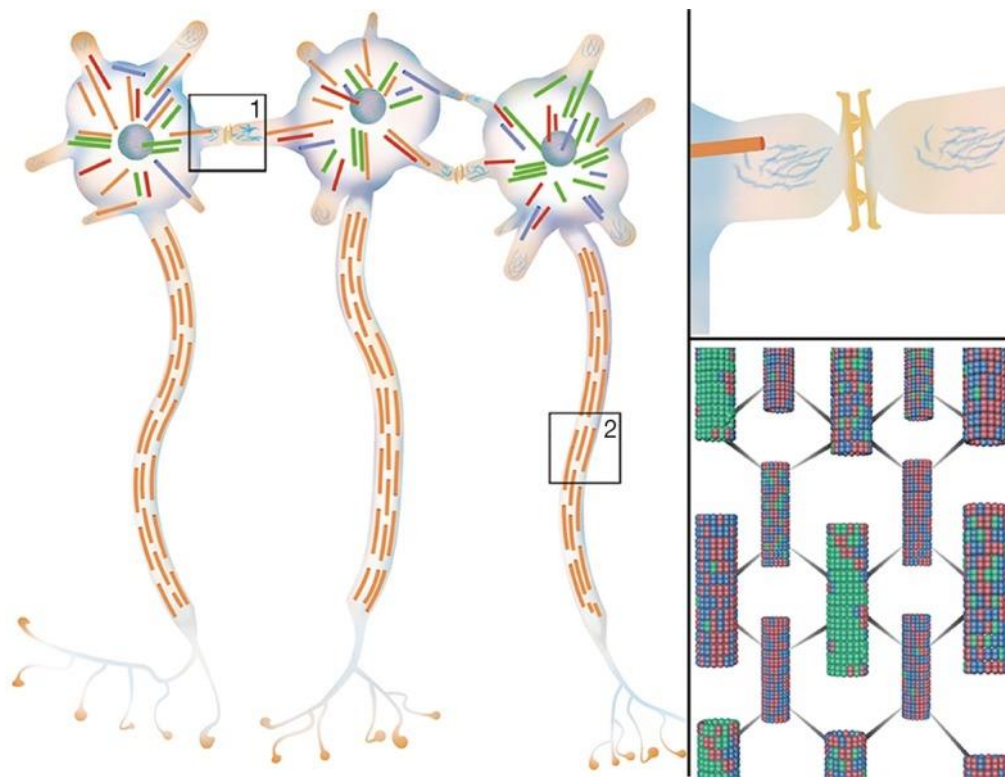


Conceptual graph of quantum space-time structure. The quantum particle vibrates from the start point O. Based on the uncertainty principle, the superposed condition of the particle determines that there are two possible tracks of particle progress in 3-dimensional xyz space. The orange and green large arrow indicates time, the fourth dimension. The initial variance between the two tracks is tiny (O point), but over time, the variance increases. One of its possible tracks would disappear after reaching the threshold at the moment of Orch-OR, leaving the other as certain and valid. The blue panel is the conscious moment of Orch-OR ([44](#)).

Originally, the cytoskeleton was proposed as the cell's nervous system and biological controller (nanocomputer), which had self-stabilizing logic algorithms introduced by Hameroff who was inspired by the subtle link between Fröhlich's coherent excitations and tubulin subunits in microtubules ([40,42,45-47](#)). He invoked that microtubules could be the fundamental units involving information processing in our enigmatic brain ([48](#)), e.g., visual identity ([49,50](#)), learning ([42](#)), cognition ([42](#)), and memory ([49,51](#)). An unprecedented collaboration heralded the advent of the Orch-OR theory. Illuminated by the structure of microtubules, Penrose posited that the microtubules composed of protein polymers played an essential role in the understanding of human consciousness from the perspective of quantum mechanics ([52](#)). Co-established by Hameroff and Penrose, the Orch-OR theory approached consciousness from various aspects and became richer and more profound with new discoveries in neuroscience and development of research methods and instruments.

Microtubules and tubulins, as the elaborate and dynamic three-dimensional network, are found in almost all eukaryotes (53,54) and some prokaryotes (55-57). Microtubules are stiff noncovalent polymers of α - and β -tubulin functioning as the cell cytoskeleton (58), the spindle during mitosis (59) and axonemes of cilia and flagella, playing dispensable roles in cell support, migration (60,61), development (62-64), gene regulation (65,66) and axoplasmic transport (67). Because of the particularity of neurons, e.g., nondividing properties and electrical signal conduction, microtubules have many specific characteristics. There are many more tubulins for microtubules within neurons than within somatic cells. The nondividing property determines its unique stability. Furthermore, each part of microtubules within neurons is also distinct. Microtubules in axons extend with the same orientation from the centrosome near the nucleus to the axon terminal. In contrast, microtubules in dendrites and cell bodies are disordered and multi-oriented (68,69) (Figure 3). The caps on dendritic and somatic microtubules could prevent treadmilling-like depolymerization that occurred in other somatic cells, which accounts for the steady and probable basis of information processing (70,71). Microtubule-associated proteins (MAPs), actin, and intermediate filaments are the key structures precipitating proper microtubule function (Figure 3). Microtubules are interconnected by MAPs; thus, their physiological activities are modulated by MAPs in different phosphorylation states (72). One of the most famous MAPs is the tau protein, which has been widely studied in Alzheimer's disease (73).

Figure 3.



Neurons with gap junctions and microtubules. Left: schematic of the morphology of adjacent neurons. Microtubules in the soma and dendrites are clustered randomly and multi-oriented, while their counterparts in axons stretch from the axon hillock to the

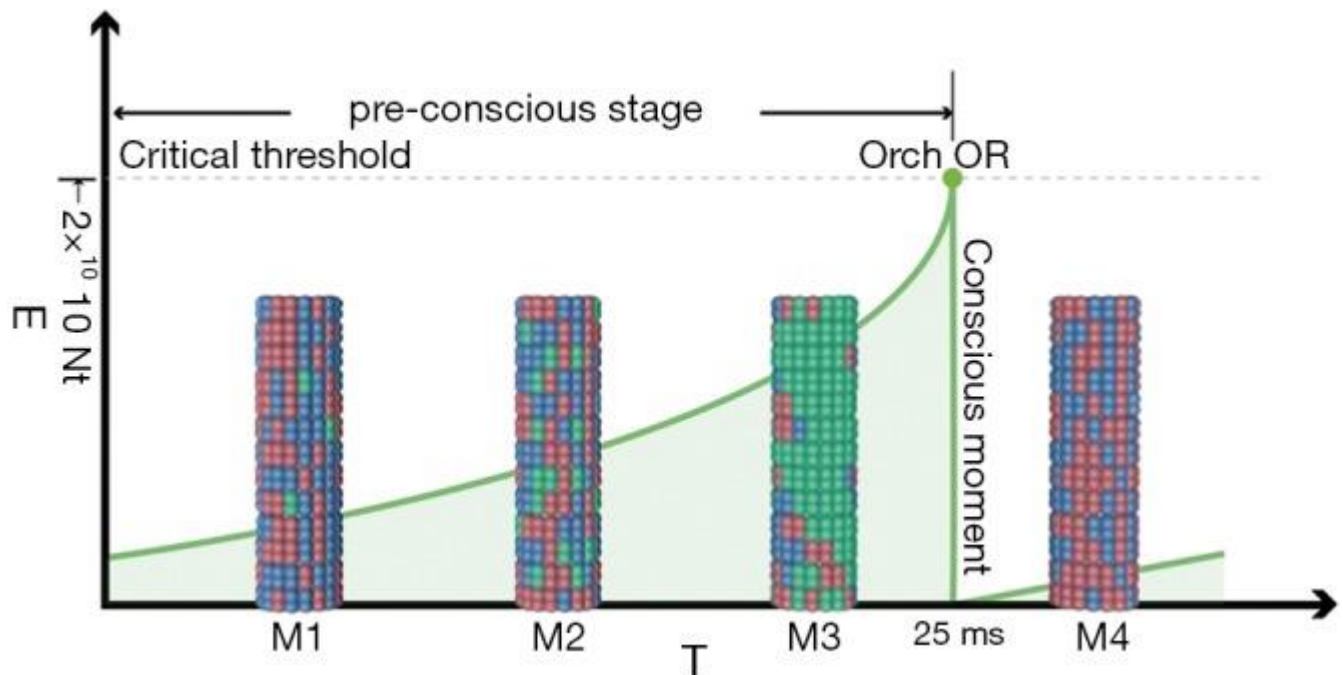
terminal in the same direction. Upper right: the magnified version of box 1 in the left graph. The gap junctions link nearby dendrites of different neurons, which make instantaneous interneuronal communication possible. Lower right: the magnified version of box 2 in the left graph. Microtubules in axons are regularly arranged in the same direction interconnected by microtubule-associated proteins (69).

Tubulins were deconstructed long ago. The study of tubulin amino acids was more advanced than the study of their α and β functional groups (74,75). In the Orch-OR theory, one of the most creative and prerequisite points is that tubulins are in a quantum superposition state according to the Heisenberg uncertainty principle; similar to qubits in quantum mechanics, they are 1 and 0 instead of 1 or 0 in computer computation.

Similar to the initiation of information transmission between neuronal synapses in traditional neurobiology, after the neurotransmitter binds to the receptor on the postsynaptic membrane, the activation of ion channel type receptors increases the intracellular calcium ion concentration. Alternatively, binding to a metabolic receptor activates a second messenger and subsequent signaling pathways. Conformational changes, e.g., phosphorylation and dephosphorylation, occur in microtubules and MAPs due to changes in calcium concentration and activation of various signaling pathways (71,76). These subtle changes caused by neurotransmitters from synaptic inputs could “orchestrate” tubulin states controlled by dipole couplings (one kind of intermolecular force), thus leading to microtubule simulation (50,77) (Figure 4). Tubulin quantum coherent superpositions and computations are increasingly combined to augment their superposed mass energy. Once the energy meets the critical threshold of quantum gravity, self-collapse occurs. That is, at this moment, the consciousness event occurs (according to Hameroff and Penrose, consciousness is discrete and independent rather than continuous) (Figure 4). Notably, under the “warm and noisy” brain, the information flow of neurotransmitters could prevent random environmental decoherence from overwhelming unitary evolution (derived from Schrödinger equation) quantum procedures. The whole orchestration and self-collapse process could be represented by a function schematic graph (Figure 4). From the equation $E = \hbar/T$ ($\hbar$ is Planck’s constant divided by 2π , T is time, E is the energy level that could be represented by the number of tubulins (N_t) in the Orch-OR theory), we could infer that only when an animal possesses enough microtubules (or a large enough brain) could it give rise to consciousness in a relatively short and realizable period of time. In the modern physical field, it was suggested that reality is composed of 3-dimensional space and 1-dimensional time. As Figure 2 shows, from the beginning, particulates (microtubules) exist two alternative states in 3-dimensional space (just like Schrödinger’s cat in the uncertainty principle). Additionally, with the orchestration of synaptic inputs over time, once the threshold is met, objective reduction occurs, and their state is confirmed (69,79). Therefore, only one certain state continued, while the other disappeared. A neuron could accept numerous distinct neurotransmitters by its dendrites and soma, that is, there must be integration between different dendritic and somatic microtubules that precipitates dendritic and somatic microtubules within a neuron to meet this critical point congruously in case their quantum coherence and computation are chaotic. Gap

junctions, which are not of much concern in ordinary neuroscience research, are considered to play an important role in quantum tunneling among dendrites in the macroscopic quantum coherent state, information exchange and mutual adjustment between neurons (49,80-82) (Figure 3). Then, the axon senses the instantaneous conscious events and fires to convey outputs to control advanced life intelligence activities and behaviors. In conclusion, dendritic and somatic microtubules' tubulins are initially in superposed states. Synaptic inputs orchestrate these tubulins such that their state tends to be unified, and total energy increases. Then, the threshold is met, followed by Orch-OR and a conscious event. Finally, all the tubulins and microtubules return to their original state, ready to accept the next synaptic input (Figure 4).

Figure 4.



Orch-OR event. Minor tubulins begin with classical computing (green tubulins in M1), which leads to quantum coherent superposition and quantum computing (expanding of green tubulins in M2 and M3). When the critical threshold of coherence to quantum gravity is met, Orch-OR will occur; thus, the entire condition of the microtubule returns to the original pattern (M4). The conscious Orch-OR event occurs in the M3 to M4 transition. The area under the curve represents superposed mass energy E with self-collapse time T , which is consistent with the formula of $E = \hbar/T$. E may be described as Nt , the number of tubulins whose mass separation (and separation of underlying space time) for time T will self-collapse. Hameroff posited that $T = 25$ ms (e.g., 40 Hz oscillations), $Nt = 2 \times 10^{10}$ tubulins (69,78). M1, microtubule state 1; M2, microtubule state 2; M3, microtubule state 3; M4, microtubule state 4.

However, debates have become increasingly turbulent since the introduction of the Orch-OR theory. Some scientists, including philosophers, physicists and

neuroscientists, support that their work heralded a new level of understanding of consciousness and was worthy of further development and experimentation ([83-88](#)). However, numerous rebuttals have emerged since the inception of the Orch-OR theory. These criticisms covered biological, gravitational and quantum fields to which Hameroff and Penrose responded one by one with detailed biological evidence, quantum mechanical equations and theoretical frameworks in “Consciousness in the universe: A review of the ‘Orch OR’ theory” ([69](#)). However, this review generated an even larger wave of criticism. Comments sprang up like mushrooms ([84,89,90](#)). One of the most noted retorts was proposed by Baars ([89](#)) in his article “Consciousness, biology and quantum hypotheses”, which caused this debate to reach the climax. Baars invoked some actual questions that many other critics mentioned. (I) All plant and animal cells are composed of microtubules, and some of their MAPs are similar. (II) Countless previous studies have proven that the cerebral cortex and thalamus are strongly implicated in conscious experience, while on a quantum basis, their specificity did not emerge. (III) More than half of the human brain region, especially the cerebellum, does not hold consciousness content; however, neurons in these regions also have identical microtubules. (IV) There is no known quantum-level proof to account for the difference between conscious versus unconscious brain events. (V) Consciousness must evolve with biological evolution. Hameroff replied with reasonable explanations that validated Orch-OR as an appropriate and qualified theory to surmise consciousness, despite the increasingly detailed development that should be interpreted ([91](#)). An inevitable situation was that arguments about Baars’s direct exclusion of “freshman” quantum theory burst into more debates in which critics stated that Baars should be more generous regarding quantum consciousness theory due to its practical basis, which could enrich our understanding of consciousness ([92-95](#)). In addition, more theories and rebuttals could bring more attention and science funding into the “puzzle game” of consciousness that has confused and captivated philosophers, neuroscientists, and physicists for ages.

Conclusion and perspective

Quantum mechanics has been inextricably linked with consciousness since its birth. Various proposals and theories have been put forward to lift the veil of consciousness. In this review, we summarized the main quantum theory of consciousness of which the theories based on neural structure were the principal parts. Undoubtedly, the Orch-OR theory co-established by theoretical physicist Penrose and neuroscientist Hameroff is currently the most convincing theory. Even more exciting, with the emergence of new drugs, new research methods, and new quantum technologies, this theory is constantly being enriched and perfected. Especially in the research of anesthesiology ([96-100](#)), memory ([71](#)), cognition ([42,101-103](#)), neural synchrony ([104](#)) and vision ([49](#)), mounting results and evidence indicated the Orch-OR theory could be self-explanatory and could be invoked to many different conscious backgrounds. More recently, Li *et al.* found that xenon’s (one kind of anesthetic) nuclear spin could impair its own anesthetic power, which involves a neural quantum process ([105](#)). Thus, the quantum theory of consciousness is increasingly gaining more supporters. With the dedication of these supporters, the quantum theory of consciousness will be gradually completed and will

be able to explain the hard problem systematically and comprehensively. As the enigmatic riddle of consciousness has remained intractable, we need more theories and hypotheses to attract enough attention and maintain lively debate. This conflict is the only way for human beings to explore the truth. Since there is no conclusive scientific mechanism of consciousness, as one of the most systemic and convincing theories among various theories of consciousness, the Orch-OR theory deserves our deeper understanding and study. Let us not put Descartes before the horse (94). It is possible to disperse the fog of our ignorance and shed light on new knowledge regarding consciousness.

Acknowledgments

We thank Yong Chen, School of Arts and Media, Tongji University for figure support.

Funding: This study was funded by Ministry of Science and Technology of China (2018YFA0107900), The National Nature Science Foundation (No. 31771491), and Shanghai Municipal Government (2019CXJQ01).

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Footnotes

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

- 1. Kuehni RG. Consciousness Explained, by Daniel C. Dennett, Little, Brown, Boston, MA, 1991, xiii + 511 pp., Color Res Appl 2010;17:295-7. 10.1002/col.5080170412 [DOI] [Google Scholar]
- 2. Marcel Kinsbourne Time and the observer: The where and when of consciousness in the brain. Behav Brain Sci 1992;15:183-201. 10.1017/S0140525X00068229 [DOI] [Google Scholar]
- 3. Koch C, Crick F. The zombie within. Nature 2001;411:893. 10.1038/35082161 [DOI] [PubMed] [Google Scholar]
- 4. Wegner DM. The Illusion of Conscious Will. Mit Pr 2003;159:197-213. [Google Scholar]
- 5. Crick F, Koch C. Some Reflections on Visual Awareness. Cold Spring Harb Symp Quant Biol 1990;55:953. 10.1101/SQB.1990.055.01.089 [DOI] [PubMed] [Google Scholar]

- 6.Koch C, Massimini M, Boly M, et al. Neural correlates of consciousness: Progress and problems. *Nat Rev Neurosci* 2016;17:307-21. 10.1038/nrn.2016.22 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 7.Brancucci A, Lugli V, Perrucci MG, et al. A frontal but not parietal neural correlate of auditory consciousness. *Brain Struct Funct* 2016;221:463-72. 10.1007/s00429-014-0918-2 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 8.Alvarez-Rodriguez U, Sanz M, Lamata L, et al. Quantum Artificial Life in an IBM Quantum Computer. *Sci Rep* 2018 4;8:14793. [[DOI](#)] [[PMC free article](#)] [[PubMed](#)]
- 9.Bravyi S, Gambetta JM, Mezzacapo A, et al. Tapering off qubits to simulate fermionic Hamiltonians. *Quantum Physics* 2017, arXiv:1701.08213.
- 10.Bravyi S, Gosset D, König R. Quantum advantage with shallow circuits. *Science* 2018;362:308-311. 10.1126/science.aar3106 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 11.Wang XL, Cai XD, Su ZE, et al. Quantum teleportation of multiple degrees of freedom of a single photon. *Nature* 2015;518:516. 10.1038/nature14246 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 12.Ding X, He Y, Duan ZC, et al. On-Demand Single Photons with High Extraction Efficiency and Near-Unity Indistinguishability from a Resonantly Driven Quantum Dot in a Micropillar. *Phys Rev Lett* 2016;116:020401. 10.1103/PhysRevLett.116.020401 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 13.Chen MC, Wu D, Su ZE, et al. Efficient Measurement of Multiparticle Entanglement with Embedding Quantum Simulator. *Phys Rev Lett* 2016;116:070502. 10.1103/PhysRevLett.116.070502 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 14.Sordillo P. Cogito Ergo Sum. *Ital Am* 2005;23:73. [[Google Scholar](#)]
- 15.Guth E. Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren. *Monatshefte Für Math Und Phys* 1932;39:A51 10.1007/BF01699179 [[DOI](#)] [[Google Scholar](#)]
- 16.Ricciardi LM, Umezawa H. Brain and physics of many-body problems. *Kybernetik* 1967;4:44-8. 10.1007/BF00292170 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 17.Stuart CI, Takahashi Y, Umezawa H. On the stability and non-local properties of memory. *J Theor Biol* 1978;71:605-18. 10.1016/0022-5193(78)90327-2 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 18. Stuart CI, Takahashi Y, Umezawa H. Mixed-system brain dynamics: Neural memory as a macroscopic ordered state. *Found Phys* 1979;9:301-27. 10.1007/BF00715185 [[DOI](#)] [[Google Scholar](#)]
- 19. Roy S, Kafatos M. Quantum processes and functional geometry: new perspectives in brain dynamics. *Forma* 2004;1:69-84. [[Google Scholar](#)]
- 20. Fröhlich H. Long-range coherence and energy storage in biological systems. *Int J Quantum Chem* 1968;2:641-9. 10.1002/qua.560020505 [[DOI](#)] [[Google Scholar](#)]
- 21. Chapin FS. Week's Citation. 1988;1989.
- 22. Marshall IN. Consciousness and Bose-Einstein condensates. *New Ideas Psychol* 1989;7:73-83. 10.1016/0732-118X(89)90038-X [[DOI](#)] [[Google Scholar](#)]
- 23. Wheeler JA. Include the Observer in the Wave Function? In: Lopes JL, Paty M. editors. *Quantum Mechanics, A Half Century Later. Episteme (A Series in the Foundational, Methodological, Philosophical, Psychological, Sociological and Political Aspects of the Sciences, Pure and Applied)*, 1977 vol 5. Springer, Dordrecht. [[Google Scholar](#)]
- 24. Bohm D. Quantum theory. *Phys Today* 1952;5:23-4. 10.1063/1.3067480 [[DOI](#)] [[Google Scholar](#)]
- 25. Gabor D. Holographic Model of Temporal Recall. *Nature* 1968;217:584. 10.1038/217584a0 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 26. Stapp HP. Mind, matter, and quantum mechanics. *Found Phys* 1982;12:363-99. 10.1007/BF00726783 [[DOI](#)] [[Google Scholar](#)]
- 27. Mahler G. *Mind, Matter and Quantum Mechanics* (2nd edition). *J Phys A Gen Phys* 2004;37:6857. [[Google Scholar](#)]
- 28. Jibu M, Yasue K. *Quantum Measurement by Quantum Brain*. 1995.
- 29. Jibu M, Yasue K. A physical picture of Umezawa's quantum brain dynamics. *Cybern Syst Res* 1992;92:797-804. [[Google Scholar](#)]
- 30. Pribram KH. *Rethinking Neural Networks: Quantum Fields and Biological Data*. Taylor & Francis, 2014. Available online: <https://books.google.com/books?id=L7NQAwAAQBAJ>
- 31. Jibu M, Yasue K. Intracellular quantum signal transfer in Umezawa's quantum brain dynamics. *Cybern Syst* 1993;24:1-7. 10.1080/01969729308961695 [[DOI](#)] [[Google Scholar](#)]

- 32.Jibu M, Yasue K. Introduction to quantum brain dynamics in, Nature, Cognition and System ed, Carvallo, M. Kluwer Academic, 1993. [[Google Scholar](#)]
- 33.Jibu M, Yasue K. Quantum Brain Dynamics and Consciousness: An introduction. John Benjamins Publishing Company, 1995. Available online: <https://books.google.com/books?id=elh1MqqymRwC>
- 34.Stenger V. The myth of quantum consciousness. Humanist 1992;53:13-5. [[Google Scholar](#)]
- 35.Chalmers DJ. Facing Up to the Problem of Consciousness. J Conscious Stud 1994;2:200-19. [[Google Scholar](#)]
- 36.Gibson KR. Book Review: The Conscious Mind: In Search of a Fundamental Theory. David J. Chalmers. Int J Quantum Chem 2015;66:107-8. [[Google Scholar](#)]
- 37.Beck F, Eccles JC. Quantum aspects of brain activity and the role of consciousness. Proc Natl Acad Sci U S A 1992;89:11357-61. 10.1073/pnas.89.23.11357 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 38.Eccles J C. Do mental events cause neural events analogously to the probability fields of quantum mechanics? Proc R Soc Lond B Biol Sci 1986;227:411-28. 10.1098/rspb.1986.0031 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 39.Amunts K, Zilles K. Architectonic Mapping of the Human Brain beyond Brodmann. Neuron 2015;88:1086-107. 10.1016/j.neuron.2015.12.001 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 40.Hameroff SR, Watt RC. Information processing in microtubules. J Theor Biol 1982;98:549-61. 10.1016/0022-5193(82)90137-0 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 41.Smith SA, Watt RC, Hameroff SR. Cellular automata in cytoskeletal lattices. Phys D Nonlinear Phenom 1984;10:168-74. 10.1016/0167-2789(84)90259-8 [[DOI](#)] [[Google Scholar](#)]
- 42.Rasmussen S, Karampurwala H, Vaidyanath R, et al. Computational connectionism within neurons: A model of cytoskeletal automata subserving neural networks. Phys D Nonlinear Phenom 1990;42:428-49. 10.1016/0167-2789(90)90093-5 [[DOI](#)] [[Google Scholar](#)]
- 43.Rasmussen S, Karampurwala H, Vaidyanath R, et al. Emergent computation in microtubule model networks. In: Century, International Conference of the IEEE Engineering, 2002. [[Google Scholar](#)]

- 44. Penrose R. The Emperor's New Mind. Neuroscience 1991;37:285-6. [\[Google Scholar\]](#)
- 45. Koruga D. Ultimate computing: Biomolecular consciousness and nanotechnology. by Stuart R. Hameroff, Elsevier Science Publishers B.V. Amsterdam, The Netherlands, 1987:357. illus. ISBN 0-444-70283-0. Biosystems 1987;22:83-4. 10.1016/0303-2647(88)90052-4 [\[DOI\]](#) [\[Google Scholar\]](#)
- 46. Hameroff SR, Rasmussen S. Information Processing in Microtubules: Biomolecular Automata and Nanocomputers. 1989.
- 47. Lahoz-Beltra R, Hameroff SR, Dayhoff JE. Cytoskeletal logic: a model for molecular computation via Boolean operations in microtubules and microtubule-associated proteins. Biosystems 1993;29:1-23. 10.1016/0303-2647(93)90078-Q [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 48. Jibu M, Hagan S, Hameroff SR, et al. Quantum optical coherence in cytoskeletal microtubules: implications for brain function. Biosystems 1994;32:195-209. 10.1016/0303-2647(94)90043-4 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 49. Woolf NJ, Hameroff SR. A quantum approach to visual consciousness. Trends Cogn Sci 2001;5:472-8. 10.1016/S1364-6613(00)01774-5 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 50. Hameroff SR. 'Funda-Mentality': Is the conscious mind subtly linked to a basic level of the universe? Trends Cogn Sci 1998;2:119-24. 10.1016/S1364-6613(98)01157-7 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 51. Hameroff SR, Craddock TJ, Tuszynski JA. "Memory bytes" - molecular match for CaMKII phosphorylation encoding of microtubule lattices. J Integr Neurosci 2010;9:253-67. 10.1142/S0219635210002482 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 52. Penrose BR. Shadows of the Mind. Oxford Univ Press, Oxford 1996;43:17-24. [\[Google Scholar\]](#)
- 53. Carlier M F, Pantaloni D. Kinetic analysis of guanosine 5'-triphosphate hydrolysis associated with tubulin polymerization. Biochemistry 1981;20:1918. 10.1021/bi00510a030 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 54. Desai A, Mitchison TJ. Microtubule polymerization dynamics. Annu Rev Cell Dev Biol 1997;13:83-117. 10.1146/annurev.cellbio.13.1.83 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

- 55.Margolin W. FtsZ and the division of prokaryotic cells and organelles. *Nat Rev Mol Cell Biol* 2005;6:862-71. 10.1038/nrm1745 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 56.Pilhofer M, Ladinsky MS, McDowell AW, et al. Microtubules in bacteria: Ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. *Plos Biol* 2011;9:e1001213. 10.1371/journal.pbio.1001213 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 57.Schlieper D, Oliva MA, Andreu JM, et al. Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc Natl Acad Sci U S A* 2005;102:9170-5. 10.1073/pnas.0502859102 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 58.Kueh HY, Mitchison TJ. Structural plasticity in actin and tubulin polymer dynamics. *Science* 2009;325:960-3. 10.1126/science.1168823 [[DOI](#)] [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
- 59.Mitchison T, Kirschner M. Dynamic instability of microtubule growth. *Nature* 1984;312:237-42. 10.1038/312237a0 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 60.Mikhailov A, Gundersen GG. Relationship between microtubule dynamics and lamellipodium formation revealed by direct imaging of microtubules in cells treated with nocodazole or taxol. *Cell Motil Cytoskeleton* 1998;41:325-40. [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 61.Watman-Storer CM, Worthylake RA, Liu BP, et al. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. *Nat Cell Biol* 1999;1:45-50. 10.1038/9018 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 62.van Eeden F, St Johnston D. The polarisation of the anterior-posterior and dorsal-ventral axes during *Drosophila* oogenesis. *Curr Opin Genet Dev* 1999;9:396-404. 10.1016/S0959-437X(99)80060-4 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 63.Bedington RS, Robertson EJ. Axis development and early asymmetry in mammals. *Cell* 1999;96:195-209. 10.1016/S0092-8674(00)80560-7 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 64.Tucker RP. The roles of microtubule-associated proteins in brain morphogenesis: a review. *Brain Res Brain Res Rev* 1990;15:101-20. 10.1016/0165-0173(90)90013-E [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 65.Christian O, Dominika I, Angela G, et al. Modulation of the expression of connective tissue growth factor by alterations of the cytoskeleton. J Biol Chem 2010;85:A42. [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 66.Rosette C, Karin M. Cytoskeletal control of gene expression: depolymerization of microtubules activates NF-kappa B. J Cell Biol 1995;128:1111-9. 10.1083/jcb.128.6.1111 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 67.Hirokawa N, Noda Y, Tanaka Y, et al. Kinesin superfamily motor proteins and intracellular transport. Nat Rev Mol Cell Biol 2009;10:682-96. 10.1038/nrm2774 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 68.Böhm KJ. P. Dustin, Microtubules (Second Totally Revised Edition). XV + 482 S., 175 Abb. Berlin-Heidelberg-New York-Tokyo 1984. Springer-Verlag. ISBN: 3-540-13283-X. J Basic Microbiol 2010;25:630. [\[Google Scholar\]](#)
- 69.Hameroff S, Penrose R. Consciousness in the universe: A review of the 'Orch OR' theory. Phys Life Rev 2014;11:39-78. 10.1016/j.plrev.2013.08.002 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 70.Guillaud L, Bosc C, Fourest-Lieuvin A, et al. STOP proteins are responsible for the high degree of microtubule stabilization observed in neuronal cells. J Cell Biol 1998;142:167-79. 10.1083/jcb.142.1.167 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 71.Craddock TJ, Tuszynski JA, Hameroff S. Cytoskeletal signaling: Is memory encoded in microtubule lattices by CaMKII phosphorylation? PLoS Comput Biol 2012;8:e1002421. 10.1371/journal.pcbi.1002421 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 72.Tuszyński JA, Hameroff S, Satarić MV, et al. Ferroelectric behavior in microtubule dipole lattices: Implications for information processing, signaling and assembly/disassembly. J Theor Biol 2015;174:371-80. 10.1006/jtbi.1995.0105 [\[DOI\]](#) [\[Google Scholar\]](#)
- 73.Brunden KR, Yao Y, Potuzak JS, et al. The characterization of microtubule-stabilizing drugs as possible therapeutic agents for Alzheimer's disease and related tauopathies. Pharmacol Res 2011;63:341-51. 10.1016/j.phrs.2010.12.002 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 74.Amos LA, Baker TS. The three-dimensional structure of tubulin protofilaments. Nature 1979;279:607-12. 10.1038/279607a0 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

- 75. Bryan RF. Biochemistry. Second Edition by Voet D, Voet JG. Acta Crystallogr 1996;52:610. [[Google Scholar](#)]
- 76. Sánchez C, Díaz-Nido J, Avila J. Phosphorylation of microtubule-associated protein 2 (MAP2) and its relevance for the regulation of the neuronal cytoskeleton function. Prog Neurobiol 2000;61:133-68. 10.1016/S0301-0082(99)00046-5 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 77. Hameroff SR, Penrose R. Orchestrated reduction of quantum coherence in brain microtubules. A Model Conscious Neural Netw World 1995;5:793-804. [[Google Scholar](#)]
- 78. Hameroff S, Nip A, Porter M, et al. Conduction pathways in microtubules, biological quantum computation, and consciousness. BioSystems 2002;64:149-68. 10.1016/S0303-2647(01)00183-6 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 79. Hameroff SR, Penrose R. Conscious Events as Orchestrated Space-Time Selections. 2015.
- 80. Hameroff S, Penrose R. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. NeuroQuantology 2007;5:1-8. 10.14704/nq.2007.5.1.114 [[DOI](#)] [[Google Scholar](#)]
- 81. Penrose R, Hameroff SR. What 'Gaps'? Reply to Grush and Churchland. J Conscious Stud 1995;2:98-111. [[Google Scholar](#)]
- 82. Hameroff S. Quantum computing in microtubules—an intra-neural correlate of consciousness? Cogn Stud 1997. doi: 10.11225/jcss.4.3_67. [[DOI](#)] [[Google Scholar](#)]
- 83. Ghosh S, Sahu S, Bandyopadhyay A. Evidence of massive global synchronization and the consciousness. Phys Life Rev 2014;11:83-4. 10.1016/j.plrev.2013.10.007 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 84. Tuszynski JA. The need for a physical basis of cognitive process. Comment on 'Consciousness in the universe: A review of the "Orch OR" theory' by Hameroff and Penrose. Phys Life Rev 2014;11:79-80. 10.1016/j.plrev.2013.10.009 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 85. Chopra D. Reality and consciousness: A view from the East. Phys Life Rev 2014;11:81-2. 10.1016/j.plrev.2013.11.001 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]
- 86. Lucas J. The face of freedom. Comment on 'Consciousness in the universe: A review of the "Orch OR" theory' by Stuart Hameroff and Roger Penrose. Phys Life Rev 2014;11:87-8. 10.1016/j.plrev.2013.11.008 [[DOI](#)] [[PubMed](#)] [[Google Scholar](#)]

- 87.Jumper CC, Scholes GD. Life—Warm, wet and noisy? Phys Life Rev 2014;11:85-6. 10.1016/j.plrev.2013.10.008 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 88.Pino S, Di Mauro E. How to conciliate Popper with Cartesius. Comment on: 'Consciousness in the universe. A review of the "Orch OR" theory' by S. Hameroff and R. Penrose. Phys Life Rev 2014;11:91-3. 10.1016/j.plrev.2013.09.001 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 89.Baars BJ, Edelman DB. Consciousness, biology and quantum hypotheses. Phys Life Rev 2012;9:285-94. 10.1016/j.plrev.2012.07.001 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 90.Tandy C. Are you (almost) a zombie?: Conscious thoughts about 'Consciousness in the universe' by Hameroff and Penrose. Phys Life Rev 2014;11:89-90; discussion 94-100. 10.1016/j.plrev.2013.10.010 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 91.Hameroff S. Quantum brain biology complements neuronal assembly approaches to consciousness: Comment on 'Consciousness, biology and quantum hypotheses' by Baars and Edelman. Phys Life Rev 2012;9:303-5. 10.1016/j.plrev.2012.07.002 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 92.Bernroider G. Is there space for the quantum domain in consciousness research?. Comment on 'Consciousness, biology, and quantum hypotheses' by Baars and Edelman. Phys Life Rev 2012;9:301-2. 10.1016/j.plrev.2012.07.004 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 93.Klimesch W, Arora A. The EEG frequency architecture, coupled oscillations and consciousness. Comment on 'Consciousness, biology and quantum hypotheses' by Baars and Edelman. Phys Life Rev 2012;9:295-6. 10.1016/j.plrev.2012.07.010 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 94.Seth AK. Putting Descartes before the horse: Quantum theories of consciousness: Comment on "Consciousness, biology, and quantum hypotheses" by Baars & Edelman. Phys Life Rev 2012;9:297-8; discussion 306-7. 10.1016/j.plrev.2012.07.005 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 95.Wessel JR. From "Neural correlates of consciousness" to "Neural causes of consciousness": A commentary on "Consciousness, biology and quantum hypotheses", by Bernard J. Baars and David E. Edelman. Phys Life Rev 2012;9:299-300; discussion 306-7. 10.1016/j.plrev.2012.07.003 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 96.Craddock TJA, Kurian P, Preto J, et al. Anesthetic Alterations of Collective Terahertz Oscillations in Tubulin Correlate with Clinical Potency: Implications

for Anesthetic Action and Post-Operative Cognitive Dysfunction. *Sci Rep* 2017;7:9877. 10.1038/s41598-017-09992-7 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

- 97. Xi J, Liu R, Asbury GR, et al. Inhalational anesthetic-binding proteins in rat neuronal membranes. *J Biol Chem* 2004;279:19628-33. 10.1074/jbc.M313864200 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 98. Pan JZ, Xi J, Tobias JW, et al. Halothane binding proteome in human brain cortex. *J Proteome Res* 2007;6:582-92. 10.1021/pr060311u [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 99. Emerson DJ, Weiser BP, Psonis J, et al. Direct modulation of microtubule stability contributes to anthracene general anesthesia. *J Am Chem Soc* 2013;135:5389. 10.1021/ja311171u [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 100. Craddock TJ, St George M, Freedman H, et al. Computational predictions of volatile anesthetic interactions with the microtubule cytoskeleton: implications for side effects of general anesthesia. *PLoS One* 2012;7:e37251. 10.1371/journal.pone.0037251 [\[DOI\]](#) [\[PMC free article\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 101. Hameroff SR. The brain is both neurocomputer and quantum computer. *Cogn Sci* 2007;31:1035-45. 10.1080/03640210701704004 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 102. Hameroff S. Quantum walks in brain microtubules - A biomolecular basis for quantum cognition? *Top Cogn Sci* 2014;6:91-7. 10.1111/tops.12068 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 103. Tonello L, Cocchi M, Gabrielli F, et al. On the possible quantum role of serotonin in consciousness. *J Integr Neurosci* 2015;14:295-308. 10.1142/S021963521550017X [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 104. Craddock TJ, Priel A, Tuszynski JA. Keeping time: Could quantum beating in microtubules be the basis for the neural synchrony related to consciousness? *J Integr Neurosci* 2014;13:293-311. 10.1142/S0219635214400019 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)
- 105. Li N, Lu D, Yang L, et al. Nuclear Spin Attenuates the Anesthetic Potency of Xenon Isotopes in Mice: Implications for the Mechanisms of Anesthesia and Consciousness. *Anesthesiology* 2018;129:271-277. 10.1097/ALN.0000000000002226 [\[DOI\]](#) [\[PubMed\]](#) [\[Google Scholar\]](#)

Consciousness as a Physical Process Caused by the Organization of Energy in the Brain

[Robert Pepperell](#)*

- FOVOLAB, Cardiff Metropolitan University, Cardiff, United Kingdom

To explain consciousness as a physical process we must acknowledge the role of energy in the brain. Energetic activity is fundamental to all physical processes and causally drives biological behavior. Recent neuroscientific evidence can be interpreted in a way that suggests consciousness is a product of the organization of energetic activity in the brain. The nature of energy itself, though, remains largely mysterious, and we do not fully understand how it contributes to brain function or consciousness. According to the principle outlined here, energy, along with forces and work, can be described as actualized differences of motion and tension. By observing physical systems, we can infer there is something it is like to undergo actualized difference from the intrinsic perspective of the system. Consciousness occurs because there is something it is like, intrinsically, to undergo a certain organization of actualized differences in the brain.

Introduction

“If mental processes are indeed physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery.”

([Nagel, 1974](#))

The philosopher Thomas Nagel summarized one of our greatest intellectual challenges: how to explain mental processes as physical processes. The aim of this paper is to outline a principle according to which consciousness could be explained as a physical process caused by the organization of energy in the brain¹.

Energy is fundamentally important in all physical processes ([Lotka, 1922](#); [Schrödinger, 1944](#); [Heisenberg, 1958](#); [Boltzmann, 1886](#)). As the biophysicist Harold [Morowitz \(1979\)](#) put it: “the flow of energy through a system acts to organize that system.” Light, chemical reactions, electricity, mechanical work, heat, and life itself can all be described in terms of energetic activity ([Chaisson, 2001](#); [Morowitz and Smith, 2007](#); [Smil, 2008](#)) as can metabolic processes in the body and brain ([Magistretti, 2008](#); [Perez Velazquez, 2009](#)). It is surprising, therefore, that energy receives relatively little attention in neuroscientific and psychological studies of consciousness. Leading scientific theories

of consciousness do not reference it ([Crick and Koch, 2003](#); [Edelman et al., 2011](#); [Dehaene, 2014](#); [Oizumi et al., 2014](#)), assign it only a marginal role ([Hameroff and Penrose, 2014](#)), or treat it as an information-theoretical quantity ([Friston, 2013](#); [Riehl et al., 2017](#)). If it is discussed, it is either as a substrate underpinning higher level emergent dynamics ([Deacon, 2013](#)) or as powering neural information processing ([Sterling and Laughlin, 2017](#)).

This lack of attention is all the more surprising given that some of the pioneers of neurobiology, psychology, and physiology found a central place for energy in their theories, including Hermann von Helmholtz (in [Cahan, 1995](#)), Gustav [Fechner \(1905\)](#), Sigmund Freud ([Gay, 1988](#)), William James ([James, 1907](#)), and Charles [Sherrington \(1940\)](#)². There are, however, signs that attention is turning again to energetic or thermodynamic-related theories of consciousness in various branches of science ([Deacon, 2013](#); [Collell and Fauquet, 2015](#); [Annala, 2016](#); [Street, 2016](#); [Tozzi et al., 2016](#); [Marchetti, 2018](#)) and in philosophy of mind ([Strawson, 2008, 2017](#)).

The present paper builds on this work by proposing that energy, and the related properties of force and work, can be described as *actualized differences* of motion and tension, and that – in Nagel’s phrase – ‘there is something it is like, intrinsically, to undergo’ actualized differences. Recent neuroscientific evidence suggests that consciousness is a product of the way energetic activity is organized in the brain. Following this evidence, I propose that we experience consciousness because there is something it is like, intrinsically, to undergo a certain organization of actualized differences in the brain.

Several researchers have tackled the problem of consciousness by treating the brain in principle as a neural information processor (e.g., [Tononi et al., 2016](#); [Dehaene et al., 2017](#); [Ruffini, 2017](#)). I will argue that the governing principle of the brain at the neural level is not information processing but energy processing. The information-theoretic approach to measuring and modeling brain activity, however, can usefully complement the energetic approach outlined here.

Consciousness and Energy in the Brain

We do not fully understand the biological function of energy in the brain or how it relates to the presence of consciousness in the person³. Given that the human brain accounts for only 2% of the body’s mass it demands a large portion of the body’s total energy budget, some 20% ([Laughlin, 2001](#); [Magistretti and Allaman, 2013](#)). Most of this energy is derived from the oxidization of glucose supplied to the cerebral tissue through the blood. Roy and Sherrington were the first to propose a direct correspondence between changes in cerebral blood flow and functional activity ([Roy and Sherrington, 1890](#)). Many features of human brain anatomy, such as the number of blood vessels per unit of space, the lengths of neural connections, the width of axons, and even the ratio of brain to stomach size are thought to be determined by the high metabolic demands associated with complex cognitive processing ([Allen, 2009](#)).

For many neuroscientists, the main function of energy in the brain is to fuel neural signaling and information processing ([Magistretti, 2013](#)); energy supply is seen as a constraint on the design and operation of the brain's computational architecture ([Laughlin, 2001](#); [Hall et al., 2012](#); [Sterling and Laughlin, 2017](#)). It has been calculated, for example, that the rate of energy supply available to the human brain places an upper 'speed limit' on neural processing of about 1 kHz ([Attwell and Gibb, 2005](#)). And [Schölvinck et al. \(2008\)](#) estimated that conscious perception of sensory stimuli increases energy consumption in primate brains by less than 6% compared to energy consumption in the absence of conscious perception⁴. They attribute this relatively small change to an energy efficient "design strategy" of the brain in which decreases in neural activity play a functional role in information processing as well as increases. Energy, on these accounts, plays no direct role in higher mental processes, like consciousness.

Robert Shulman and colleagues have argued there is a direct connection between energy in the brain and consciousness ([Shulman et al., 2009](#); [Shulman, 2013](#)). By studying the progressive loss of behavioral response to external stimulus from wakefulness to deep anesthesia, they found a corresponding reduction and localization of cerebral metabolism (a marker of energy consumption). Therefore, they argue, high global metabolism is necessary for consciousness. However, they are also clear that high global metabolic rates are not sufficient as patients with locked-in-syndrome and those who suffer from some forms of epileptic seizure can register high levels of global brain metabolism without exhibiting the observable behavior that we expect from a conscious person ([Shulman, 2013](#); [Bazzigaluppi et al., 2017](#)). Shulman's thesis has been challenged on several grounds ([Seth, 2014](#)). For example, it has been pointed out that behavioral responsiveness may be inadequate as a measure of sentience given that vestiges of consciousness have been detected in people diagnosed as being in a vegetative state with a low cerebral metabolism ([Owen et al., 2006](#)). Moreover, some patients who recover from a vegetative state to regain consciousness do so despite having substantially reduced cerebral metabolism compared with normal controls ([Laureys et al., 1999](#); [Chatelle et al., 2011](#)).

In recent years there has been a growing interest in intrinsic brain activity ([Clarke and Sokoloff, 1999](#); [Raichle, 2011](#)). This background or spontaneous activity occurs in the resting awake state in the absence of external stimulation or directed attention, and its energy demands can greatly exceed those of localized activation due to task performance or attention. The discovery of this so-called 'dark energy' in the brain ([Raichle, 2010](#)) was greeted with some surprise in the neuroscience community and remains controversial ([Morcom and Fletcher, 2007](#)). Work on intrinsic activity led to the identification of a 'default mode network' in the brain, an extended set of interconnected regions that uses high levels of energy when a person is in a non-attentive state. Energy use drops significantly in this network when a more cognitively demanding task, such as paying attention to a stimulus, is performed ([Shulman et al., 1997](#); [Raichle et al., 2001](#)). [Vanhaudenhuyse et al. \(2009\)](#) reported that connectivity within the default mode network in patients with severe brain-damage deteriorates in proportion to the degree of conscious impairment, suggesting it plays an important role in sustaining consciousness.

Meanwhile, it is somewhat surprising to find that energy use during non-rapid eye movement sleep remains at ~85% of that in the waking state, while during rapid eye movement sleep it can be as high as in the waking state ([Dinuzzo and Nedergaard, 2017](#)). At the same time, consciousness can be minimally sustained with energy use at only 42% of the level that occurs in healthy conscious individuals, suggesting that much cerebral metabolic activity in normal waking states does not directly contribute to consciousness ([Stender et al., 2016](#)). Many anesthetic agents are thought to obliterate consciousness because they reduce the global rate of cerebral metabolism ([Hudetz, 2012](#)). Administering ketamine, on the other hand, increases brain metabolism yet can still lead to loss of responsiveness ([Pai and Heining, 2007](#)). Overall, it seems we find no clear correlation between the total amount of energy used by the brain, or the location where the energy is used, and the level of consciousness detectable in the person.

Consciousness and the Organization of Energetic Processing in the Brain

An alternative, or perhaps complementary, way to think about this issue is in terms of how the energetic activity in the brain is organized rather than its global level or localization. Indeed, this has implicitly been the focus of recent research that aims to provide quantitative measures of consciousness levels. In one study, researchers used transcranial magnetic stimulation (TMS) to send a magnetic pulse through the brains of healthy controls and patients with various states of impaired consciousness ([Casali et al., 2013](#)). By measuring how the pulse perturbed the cortex the researchers were able to determine the relative complexity and extent of the pathways through which the pulse propagated and correlate these to levels of consciousness. The researchers calculated a perturbation-complexity index (PCI) that quantified the levels of consciousness present in each person they studied. This method was further validated as a reliable objective measure of levels of consciousness by [Casarotto et al. \(2016\)](#).

The PCI was calculated using data from electroencephalographic (EEG) measurements of the cerebral perturbation following the TMS. Images from the EEG were filtered into binary data that was then analyzed using a Lempel–Ziv algorithm, a commonly used information-theoretical technique in which complexity is measured as a function of data string compressibility, with more complex data strings being less compressible ([Ziv and Lempel, 1977](#); [Aboy et al., 2006](#)). Other researchers have developed similar information-theoretical methods for quantifying the complexity of brain activity and levels of consciousness. [King et al. \(2013\)](#) analyzed data from 181 EEG recordings of patients who were diagnosed with varying states of impaired consciousness and applied a measure of weighted symbolic mutual information (wSMI) that sharply distinguished between patients in vegetative state, minimally conscious state, and conscious state.

Although information theoretic tools were being used to analyze and interpret the data in these studies we should note that what was actually being detected by the experimental procedures was not information *per se* but the organization of energetic activity or processing in the brain. Energetic processing – the processes by which the brain regulates the flow of energy through its structures – is routinely detected at varying

degrees of spatial and temporal resolution, either directly or indirectly, by neuroimaging techniques such as positron emission tomography (PET), functional magnetic resonance image (fMRI), and EEG ([Niedermeyer and Lopes da Silva, 1987](#); [Bailey et al., 2005](#); [Shulman, 2013](#)). Referring again to the study by [Casali et al. \(2013\)](#), the perturbations from which the PCI was calculated were generated by a pulse of magnetic energy from the TMS and were imaged with EEG that measures electrical voltage differences, that is, fluctuations in energetic potentials between clusters of neurons in the cortex ([Niedermeyer and Lopes da Silva, 1987](#); [Hu et al., 2009](#); [Koponen et al., 2015](#)). The PCI and wSMI can therefore be interpreted as measures of the complexity or organization of energetic processing in the brain during the experimental procedures.

Subsequent research has directly investigated the connection between brain metabolism (how the brain regulates energy conversion), brain organization, and levels of consciousness by combining EEG measures with PET, a more specific measure of cerebral metabolism. [Chennu et al. \(2017\)](#) collected data from 104 patients in varying states of conscious impairment using both techniques. By analyzing this data, they determined a metric that discriminated levels of consciousness to a high degree of accuracy. This study built on previous work by [Demertzi et al. \(2015\)](#) that used fMRI to correlate a measure of intrinsic functional connectivity in the brain with levels of consciousness. The PCI method has been further validated by a study combining EEG and ¹⁸F-fluorodeoxyglucose (FDG)-PET ([Bodart et al., 2017](#)), so reinforcing the link between levels of consciousness and the organization of metabolic activity in the brain.

Current brain imaging methods do not strictly speaking detect information processing⁵. They do, however, detect changes associated with increases in energy consumption (via fMRI and PET) and fluctuations in electrical potential energy (via EEG), both of which reliably correlate with changes in mental function and behavior. On the basis of what we can observe, the brain operates according to the principle of energetic processing. The evidence discussed above suggests levels of consciousness are determined by the organization of energy processing in the brain rather than on its global level or localization; wakeful conscious states are associated with more complex organization. To understand why this might be we need to consider the concept of energy in more depth.

Energy

The concept of energy that we are familiar with today emerged only slowly from its beginnings in the late eighteenth century. It developed through the study of thermodynamics in the nineteenth century, and then found its place at the center of theories of relativity, quantum mechanics, and cosmology in the twentieth ([Coopersmith, 2010](#)). In colloquial usage energy refers to ideas of vigor, vitality, power, activity, and zest. In scientific usage, however, energy is defined as the ability of a system to do work⁶. Work is defined as the transfer of energy involved in moving an object over a distance by an external force, at least part of which is applied in the direction of the displacement ([Duncan, 2002](#)). Scientists and engineers often refer to energy as an abstract property: “Energy is a mathematical abstraction that has no existence apart

from its functional relationship to other variables” ([Abbott and Van Ness, 1972](#); [Rose, 1986](#)). It is a property that can be converted from one form to another, and in an isolated system the total quantity is conserved ([Smil, 2008](#)).

Despite the enormous amount of interest in the physics of energy and its central importance in so many branches of science, its nature remains in many ways mysterious ([Feynman, 1963](#); [Smil, 2008](#); [Coopersmith, 2010](#)) and it has been the subject of relatively little philosophical interrogation ([Coelho, 2009](#)). Treating energy as an abstract accounting quantity is perfectly satisfactory for many scientific purposes, where there is little reason to question its nature. But if energetic activity plays a significant role in consciousness, as the evidence cited above suggests it might, then its nature deserves closer scrutiny.

The concept of energy in the European intellectual tradition can be traced back to Aristotle who used but never precisely defined the term *energeia* (ἐνέργεια) to convey the ideas of action, activity, actuality, being at work, and acting purposefully (Sachs in [Aristotle, 2002](#)). Scholars have long debated the best way to translate *energeia* from ancient Greek. The word ‘energy’ itself has been used, as have ‘activity’ and ‘actuality,’ but ‘being-at-work’ is currently favored, partly due to *energeia*’s roots in *ergon*, the ancient Greek for work ([Aristotle, 1818](#); [Ellrod, 1982](#); Sachs in [Aristotle, 2002](#)). Modern scholars have tended to quarantine the ancient concept of *energeia* from contemporary ideas about energy. But Aristotle’s term may still have value when thinking about energy’s nature. This is especially so when we take into account the ideas of Aristotle’s intellectual forebear Heraclitus, whose cosmological view was informed by three main principles: (i) that activity in nature is driven by ‘fire’ – which has been interpreted as synonymous with energy ([Heisenberg, 1958](#)), (ii) is subject to continual flux or motion, and (iii) is structured by antagonism or tension and ([Kahn, 1989](#); Sachs in [Aristotle, 2002](#)).

We now understand there to be two main forms of energy: kinetic and potential. Kinetic energy is possessed by objects due to their motion, while potential energy is possessed by objects due to their relative position or configuration. All other forms of energy, such as thermal, electromagnetic, solar, chemical, gravitational, atomic, and so on are in themselves forms of either kinetic or potential energy ([Duncan, 2002](#); [Smil, 2008](#)). Much can be said about kinetic and potential energy, including the fact that they are causally efficacious, that is, they cause real change and activity in the material world⁷. But I want to draw attention here to the fact that they are both manifestations of *difference*. Kinetic energy is difference as motion or change; potential energy is difference as tension or antagonism. Neither kinetic nor potential energy inhere absolutely in objects but are relational properties; motion or change occurs relative to a frame of reference, and tension or antagonism occurs between one object, or force, and another. The concept of *difference* then is of utmost importance when considering the nature of energy and the related properties of force and work⁸.

Actualized Difference

If energy is the ability to do work then the displacement of a body undergoing work is due to force, defined as the “agency that tends to change the momentum of a massive body” ([Rennie, 2015](#)) or less formally as a “push or a pull.” Forces act and react antagonistically in equally opposing pairs and are therefore, like energy, manifestations of difference. The discipline of physics finds it convenient to treat energy, forces and work as distinct quantities to be balanced in abstract mathematical equations. But in nature they are integral and actualized, acting collectively in time and space with causal efficacy.

By observing nature, we can infer there is ‘something it is like’ to be a physical system undergoing antagonistic forceful interactions, and what it is like will vary as the interactions vary⁹. There is something it is like, for example, to be a piece of rope undergoing great tension that is different from what it is like to be the same rope when relaxed, or for a rock to crash to earth having been in freefall. Some effects of these interactions may be observed from an extrinsic perspective; we may hear a creak or a crunch. But the something it is like to undergo the interactions themselves is an intrinsic property of the observed system to which the extrinsic observer has no access. It is for this reason that its presence and nature can only be inferred¹⁰.

This is not to claim that forces acting at the subatomic scale between particles, or at the macroscopic scale in rope or rock, undergo anything like the experience we undergo as conscious humans¹¹. Something it is like-ness is not in itself sufficient for consciousness. Rather it is to recognize that:

- (i) energy, forces, and work are *actualized*,
- (ii) they are expressions of *difference*, and
- (iii) there is *something it is like*, intrinsically, to undergo actualized difference.

I use the term *actualized difference* to refer to the active, antagonistic nature of energy, forces and work in a way that encompasses Heraclitean cosmology, Aristotelian *energeia*, and contemporary scientific descriptions of energy. Mathematical equations can represent actualized differences with abstract differences, in the form of symbols and numbers, but not whatever it is that puts the “fire in the equation” ([Hawking, 1988](#))¹². For that we must refer back to nature itself – to the territory rather than the map ([Korzybski, 1933](#)).

Energy and Information

For many contemporary scientists, information is a fundamental property of nature. For some it is *the* most fundamental property of nature ([Davies, 2010](#)). Neuroscientists often claim that the brain operates according to the principle of information processing. We read that “the brain is fundamentally an organ that manipulates information” ([Sterling and Laughlin, 2017](#)) and that brains are “information processing machines” ([Ruffini, 2017](#)). Individual neurons are treated as information processing units, while neural firing patterns are converted into sequences of binary digits (1s and 0s) that encode information ([Koch, 2004](#)). Recent prominent theories claim consciousness is identical

with ([Tononi et al., 2016](#)) or results from ([Dehaene et al., 2017](#)) certain kinds of information structures or information processes in brains.

Information is variously and sometimes imprecisely defined in science ([Capurro and Hjørland, 2003](#)), its meaning is still strongly contested ([Lombardi et al., 2016](#); [Roederer, 2016](#)), and many people regard it as being to some extent subjective, relativistic, or observer-dependent ([von Foerster, 2003](#); [Deacon, 2010](#); [Werner, 2011](#); [Logan, 2012](#); [Searle, 2013](#); [de-Wit et al., 2016](#)). The term is often used in science colloquially (meaning ‘what is conveyed by an arrangement of things’) or “intuitively” ([Erra et al., 2016](#)). And where one might expect to find a clear definition, such as in a dictionary of physics, biology or chemistry, none appears ([Hine, 2015](#); [Rennie, 2015, 2016](#)).

The most widely cited technical definition of information is that given by [Shannon \(1948\)](#) as part of his mathematical theory of communication. For Shannon, information does not refer to meaning or semantics, as it does colloquially. The information is the amount of uncertainty in a message (a sequence of data) measured through probabilistic analysis of its elements. Information theory has developed into an exceptionally powerful mathematical tool that can be used, among many other things, to measure the complexity of physical systems. But a quantity of Shannon information is a measure of what can be *known* about a system as distinct from the system itself. The information lies with the measurer rather than the measured¹³.

The other commonly cited definition of information is Gregory Bateson’s “a difference that makes a difference” ([Bateson, 1979](#)). Like his fellow cybernetic theorist [Wiener \(1948\)](#), Bateson sharply distinguished information from energy. Difference is not a property of what he calls the “ordinary material universe” governed by energetic activity. It is not subject to the effects of impacts and forces but is an abstract, relational property of the mind that exists outside the realm of physical causation: “Difference, being of the nature of relationship, is not located in time or space.” Information defined according to Bateson as a “non-substantial” abstract difference cannot be used to explain consciousness as a physical process¹⁴.

The integrated information theory of consciousness (IIT) proposed by [Tononi et al. \(2016\)](#) provides an alternative, non-Shannonian, definition of information as “a form in cause-effect space.” Cause-effect space, according to their theory, contains a “conceptual structure” – a constellation of related concepts – that is specified by the “physical substrate of consciousness” (PSC), this being the precise complexes of neural activation involved in any experience. Each conscious experience is identical with this “form,” denoted $\Phi^{\max}$ when maximally integrated. But while IIT is presented as a theory of integrated information, it could equally serve as a theory of how energetic processing is organized since the PSC consists in the causally interrelated patterns of neural firing that are identical with the conscious experience.

Treating brains as neural information processors does not help us to understand consciousness as a physical process because information, according to the commonly accepted definitions, is not a physical property of brains at the neural level; *there is no information in a neuron*¹⁵. It is useful, however, to apply information-theoretical methods

to study the organization of physical systems, such as brains. [Wiener \(1948\)](#) stated: "...the amount of information in a system is a measure of its degree of organization..." As exemplified in several studies and theories cited here, we can measure and model the way the organization of energetic processes in the brain contributes to the presence of consciousness in a person¹⁶. But the abstract difference between 0 and 1 is not equivalent to the actualized difference between a neuron at rest and firing.

The Brain as a 'Difference Engine'

The challenge of explaining consciousness as a physical process is made more tractable, I suggest, by recognizing that brains operate on the principle of energetic processing. Neurons, in concert with other material structures such as astrocytes and mitochondria, convert, distribute, and dissipate electro-chemical energy through highly organized pathways in order to drive behaviors critical to the organism's survival. This makes sense when we consider the fact that organisms inhabit a physical world that is structured through the actions of energy, forces and work. To survive and prosper in this world they must continually work to acquire new supplies of high-grade or free energy to maintain an internal state far from thermodynamic equilibrium ([Boltzmann, 1886](#); [Schrödinger, 1944](#); [Schneider and Sagan, 2005](#)). Besides internal regulation, nervous systems enable organisms to perform two major tasks: *discriminating* between variations in environmental conditions, such as temperature, acidity, salinity, nutrient levels, or presence of predators, and *moving* toward environmental conditions that are beneficial to survival and away from those that are harmful.

The mechanisms that enable performance of these tasks can be seen at work in organisms with relatively simple nervous systems, such as the *C. elegans* worm ([Sterling and Laughlin, 2017](#)). Chemical gradients in the environment activate chemosensory neurons on the worm's surface that connect via interneurons to motor neurons that control the action of dorsal and ventral muscles, which, in turn, control the worm's movement ([de Bono and Maricq, 2005](#)). In this way, differences of chemical potential energy in the environment are converted into differences of electro-chemical energy in the sensing apparatus of the organism and then into differences of chemical energy in the muscles, which, by antagonistic action, are converted into the kinetic energy of the organism's movement. The organism makes discriminations in the environment relevant to its interests so that it can take appropriate actions in response.

We can see the same basic principle at work in biology of far greater complexity. The human visual system, for example, is highly demanding on the brain's energy budget ([Wong-Riley, 2010](#)). But the evolutionary benefit of human vision is the capacity it confers to guide finely controlled bodily actions in light of environmental conditions. This is achieved through an intricate sequence of energy conversions, beginning with the arrival at the retina of electro-magnetic energy from the environment and cascading through numerous energetic exchanges in the neural pathways of the visual system that progressively differentiate features of the environment ([Hubel and Livingstone, 1987](#)). This frequently results in the conversion of electro-chemical energy in the motor system and muscles to the kinetic energy of bodily movement ([Goodale, 2014](#)). The fact that

our complex biology supports so rich a repertoire of sensory discriminations and motor responses may be only a difference of degree rather than of kind with the humble worm.

We might think of sensory cells responding to stimulation from environmental energy by becoming excited or by increasing local neural activation. But vertebrate photoreceptors are, contrary to what one might expect, hyperpolarized by photon absorption. This means they ‘turn off’ when exposed to light and ‘turn on’ in the dark, even though they use more energy in the dark ([Wong-Riley, 2010](#))¹⁷. Meanwhile, some of the neurobiological evidence cited in Section “Consciousness and Energy in the Brain” cautions us against assuming that sensory stimulation always results in increased neural activation. Decreases in activation in the brain can occur in response to cognitively demanding tasks, yet can go unnoticed in imaging studies with methodologies designed to detect task-evoked increases in metabolic rate above baseline ([Raichle et al., 2001](#); [Schölvinck et al., 2008](#)). And of course not all neural activation is excitatory; neural inhibition is vitally important in brain function, as elsewhere in the nervous system, and this also entails an energetic cost ([Buzsáki et al., 2007](#)). There is evidence that an optimal balance between neural excitation and inhibition (E-I balance) in the cerebral cortex is critical for the brain to function well ([Zhou and Yu, 2018](#)).

In light of these mechanisms, the energy-hungry brain might be understood as a kind of ‘difference engine’ that works by actuating complex patterns of motion (action potential propagation) and tension (antagonistic pushes and pulls between forces) at various spatiotemporal scales. Firing rates and electrical potentials vary within neurons, between neurons, between networks of neurons, and between brain regions, so maximizing the differential states the brain undergoes. A decrease in activation, or a reduction in firing rate, can create a differential state just as much as an increase. And, as is indicated by the work of [Schölvinck et al. \(2008\)](#), deactivation may be an energy efficient way for the brain to increase its repertoire of differential states. Maintaining a global E-I balance across spatiotemporal scales, meanwhile, is thought to promote ‘efficient coding’ in sensory and cognitive processing ([Zhou and Yu, 2018](#)). All this lends support to the idea, proposed above, that one of the roles of energetic activity in the brain is to efficiently actuate *differences* of motion and tension that advance the interests of the brain-bearing organism. It is the *actualized difference* that makes the difference.

Energetic Organization as the Cause of Consciousness

In theory, we could account for all the highly complex processes occurring in the brain in terms of energy, forces and work, that is, as physical, chemical and biological processes. But the seemingly unassailable problem of how any of these processes might cause consciousness remains. The principle outlined here – that there is something it is like, intrinsically to undergo differences due to the antagonistic action of energy, forces and work – may offer a toehold in the slippery face of the problem. There is something it is like, intrinsically, to be a tense muscle that is different from being a relaxed muscle. There is something it is like, intrinsically, to be networks of neurons in

fantastically complex states of actualized differentiation from other networks, with action potentials propagating through vast arrays of fibers. But all this something is it like-ness is not in itself sufficient for consciousness. Muscles are not conscious, and networks of neurons are active in the brain when we are in dreamless sleep or under anesthesia. What is it about the organization of energetic processes in the brain, as discussed in Section “Consciousness and the Organization of Energetic Processing in the Brain,” which determines the level of consciousness we experience?

We gain some insight into the association between consciousness and the organization of energetic processing in the brain from studies of anesthesia. The reason why anesthetic agents obliterate consciousness is not understood ([Mashour, 2004](#)). Recent work has focused on the ways in which they interfere with the brain’s capacity to generate patterns of localized differentiation (often termed ‘information’) and to bind together or integrate those patterns across widely distributed brain networks ([Hudetz, 2012](#); [Hudetz and Mashour, 2016](#)). Evidence from studies on the neurological effects of anesthetics suggests that consciousness is lost as distant regions of the brain become functionally isolated and global integration breaks down ([Lewis et al., 2012](#)). The idea that consciousness depends on maximizing differentiation and integration in the brain lies at the heart of IIT ([Tononi, 2012](#); [Oizumi et al., 2014](#)).

A potential mechanism supporting global integration of local differentiation is recurrent or reentrant processing, in which widely distributed areas of the brain engage in complex loops of cortical feedback via massively parallel connections ([Edelman et al., 2011](#); [Edelman and Gally, 2013](#)). A number of studies of the effects of anesthetics have shown that they disrupt feedback connectivity, and hence integration, particularly in the frontoparietal area of the brain ([Lee et al., 2009](#); [Hudetz and Mashour, 2016](#)). Studies of brain organization during deep sleep have also reported an increase in modularity consistent with the loss of integration among regions of the brain found in the awake state ([Tagliazucchi et al., 2013](#)). This suggests that the presence of consciousness in a wakeful person depends on a certain level of functional integration supported by cortical feedback loops ([Edelman, 2004](#); [Alkire et al., 2008](#)) but it is not known how or why.

A major contribution of cybernetic theory was to recognize the importance of feedback mechanisms for controlling behavior in mechanical and living systems ([Wiener, 1948](#); [Bateson, 1972](#)). Feedback systems are self-referential; one part of the system casually affects another, which in turn affects the first. Such systems are apt to generate behaviors that are an irreducible property of the system as a whole ([Hofstadter, 2007](#); [Deacon, 2013](#)). One example is video feedback, which occurs when a video camera is pointed at a monitor showing the output from the camera ([Crutchfield, 1984](#)). When correctly arranged the monitor will at first show a tunnel-like image that will then spontaneously ‘blossom’ into an intricate, semi-stable pattern of remarkable diversity and fascinating beauty (see Figure 1)¹⁸. Since this is an energetically actuated process we can infer, following the arguments already given, that there is something it is like to be the video feedback system in full bloom, from its intrinsic perspective. But it is not conscious.

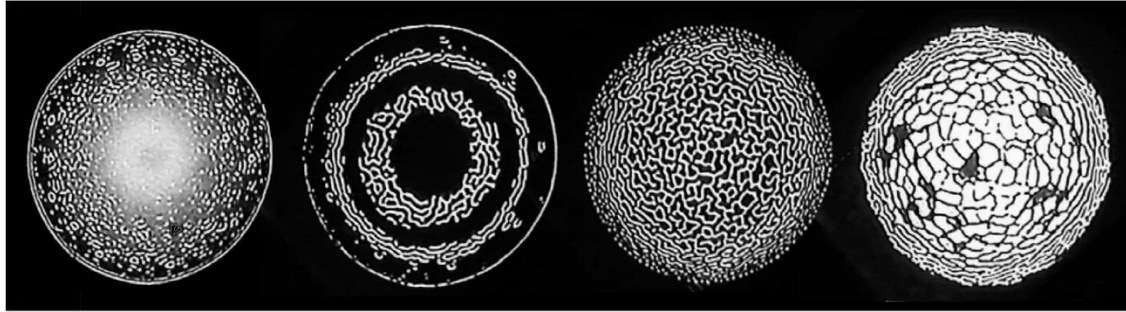


FIGURE 1. Stills from a video feedback sequence generated by the author. These patterns are created by pointing a video camera at a screen showing the camera's output. What begins as tunnel-like images soon 'blossoms' into an ever-changing pattern of great diversity and fascinating beauty. [Pepperell \(2018\)](#).

Gerald Edelman proposed that “phenomenal experience itself is entailed by appropriate reentrant intracortical activity” ([Edelman and Gally, 2013](#)). The human brain undergoes recursive or reentrant behavior of an unimaginably higher order of complexity than in the video system¹⁹. But the underlying operating principle may be analogous. Video feedback arises because the system is organized as a self-observing loop. If we assume that reentrant activity in the brain is also a kind of self-observing loop in which processes in one part the brain both affect and are affected by processes in other parts, then we can envisage a kind of pattern blooming in the brain analogous to that we see in video feedback. This pattern would be actuated by sufficiently organized electro-chemical activity, among neurons and neurotransmitters, channeled through reentrant neural circuits.

The something it is like-ness a brain organized in this way would be undergoing is of a different order to that of a brain with diminished integration in dreamless sleep or under anesthesia. No other physical system, as far as we know, has the same capacity for complex (differentiated and integrated) recursive processing as the human brain, and that dynamic organization reaches its apogee when we are wakefully conscious, as suggested by the evidence cited in Section “Consciousness and the Organization of Energetic Processing in the Brain.” When the energetic processes in our brains are operating at a certain level of dynamic recursive organization – the “appropriate reentrant intracortical activity” – then we undergo something it is like, intrinsically, to undergo something it is like, intrinsically, to undergo something it is like ... recursively. In other words, *there is something it is like, intrinsically, to be something it is like, recursively, to undergo the particular organization of actualized differences found in the conscious brain*. For this, we have the most direct and irrefutable evidence possible – what it's like to undergo our own conscious experience²⁰.

Is it reasonable then to propose that consciousness is *caused* by the way energetic activity is dynamically and recursively organized in the brain? It is no less reasonable than attributing the causes of other biological phenomena, such as the behavior of the nematode worm, to the way energetic activity is organized. If consciousness is a physical (biological and chemical) process, and if physical processes are caused by energetic activity (alongside forces and work), then consciousness, in principle, could be caused by energetic activity and the way it is organized.

Naturalizing Consciousness

In 1937–8, Charles Sherrington gave a series of lectures on the relationship between energy and mind, collected in the volume *Man on his Nature* ([Sherrington, 1940](#)). In line with the physics of his day, Sherrington understood the natural world to be composed of forms of energy. But he could not conceive how the mind could be forged from energy: “The energy-concept of Science collects all so-called ‘forms’ of energy into a flock and looks in vain for the mind among them.” The mystery was deepened for him by the knowledge, then emerging through studies of electrical and metabolic activity in the brain, of how intimately energy and mental function must be linked. He was compelled to wonder “Is the mind in any strict sense energy?” but reluctantly concluded that “...thoughts, feelings, and so on are not amenable to the energy (matter) concept.” They lie beyond the purview of natural science, despite the “embarrassment” this causes for biology.

If we are to naturalize consciousness, then we must reconcile energy and the mind. I have outlined a principle that may help to explain consciousness as a physical process. It entails re-examining the modern scientific concept of energy in the light of Aristotle’s *energeia* and its Heraclitean roots. Accordingly, we arrive at a view of physical processes in nature as actualized differences of motion and tension. Sherrington understood that “Energy acts, i.e., is motion.” But he went on “...of a mind a difficulty is to know whether it is motion.” Treating the brain as a difference engine that works to actualize and organize differences of motion and tension to serve the interests of the organism is, I submit, a natural approach to understanding consciousness as a physical process.

Conclusion

If consciousness is a natural physical process then it should be explicable in terms of energy, forces and work. Energy is a physical property of nature that is causally efficacious and, like forces and work, can be conceived as actualized differences of motion and tension.

Evidence from neurobiology indicates that the brain operates on the principle of energetic processing and that a certain organization of energy in the brain, measured with information theoretic techniques, can be reliably predict the presence and level of consciousness. Since energy is causally efficacious in physical systems, it is reasonable to claim that consciousness is in principle caused by energetic activity and how it is dynamically organized in the brain.

Information in the biological context is best understood as a measure of the way energetic activity is organized, that is, its complexity or degree of differentiation and integration. Information theoretic techniques provide powerful tools for measuring, modeling, and mapping the organization of energetic processes, but we should not confuse the map with the territory.

Actualized differences, as distinct from abstract differences represented in mathematics and information theory, are characterized by there being something it is like, intrinsically, to undergo those differences, that is, to undergo antagonistic states of opposing forces. All actualized differences undergo this something it is like-ness, but not all contribute to consciousness.

It is proposed that a particular kind of activity occurs in human brains that causes our conscious experience. It is a certain dynamic organization of energetic processes having a high degree of differentiation and integration. This organization is recursively self-referential and results in a pattern of energetic activity that blossoms to a degree of complexity sufficient for consciousness.

If consciousness is a physical process, and physical processes are driven by actualized differences of motion and tension, then there is something it is like to undergo actualized differences organized in a certain way in the brain, and this is what we experience – intrinsically²¹.

Author Contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

I wish to acknowledge the support of the Vice Chancellor's Board of Cardiff Metropolitan University.

Conflict of Interest Statement

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

I am grateful to the following for discussions and suggestions: Alistair Burleigh, Alan Dix, Chris Doran, Robert K. Logan, Heddwyn Loudon, Chen Song, Galen Strawson, Emmett Thompson, Sander Van der Cruys, and to the reviewers for their comments and criticism.

Footnotes

1. [△] I take it that physical processes occur in time and space and are causally determined by the actions of energy, forces and work upon matter. I take

consciousness to be the capacity for awareness of self and world, which is particularly highly developed in humans.

2. [^] For further discussion on the historical context see [Pepperell \(2018\)](#).
3. [^] Although for the sake of brevity I refer in this paper to consciousness occurring in the brain, consciousness is something that people undergo. Brains cannot sustain consciousness independently of the people in which they are housed ([Pepperell, 1995](#)).
4. [^] Strictly speaking energy is not consumed but converted from one form to another.
5. [^] The authors of [Wollstadt et al. \(2017\)](#), for example, studied the breakdown of local information processing under anesthesia using information theoretic methods. They point out that the EEG procedure they used did not directly record information processing in the brain but local field potentials, that is, fluctuations in quantities of potential energy.
6. [^] There seems to be an ambiguity in some textbooks about whether energy is an enabling property possessed by a system or body, e.g., [Duncan \(2002\)](#), or a measure of such a property, e.g., [Rennie \(2015\)](#). I will take energy to be a property possessed by systems or bodies, quantities of which can be measured.
7. [^] “Energy may be called the fundamental cause for all change in the world” ([Heisenberg, 1958](#)). The neurobiologist Gerald Edelman neatly defined causal efficacy as “The action in the physical world of forces or energies that lead to effects or physical outcomes” ([Edelman, 2004](#)).
8. [^] Neuroanthropologist Terence Deacon defines energy as a “relationship of difference” ([Deacon, 2013](#)). Note that energy is difference, but not all differences are energy; red is a color, not all colors are red.
9. [^] Nagel clarified the term ‘something it is like’ as meaning not what something resembles but ‘how it is’ for the system ([Nagel, 1974](#)).
10. [^] Note that this claim is not as far-fetched as it might at first seem: If (i) consciousness in people is a physical process – due to energy, forces and work – and (ii) we infer the presence of consciousness in other people on the basis of observing them extrinsically – as we habitually do – and (iii) there is something it is like to be a conscious person – as we assume there is – then (iv) we routinely infer the presence of an intrinsic something it is like-ness in a physical process on the basis of observing it from an extrinsic perspective. However, as discussed below, human consciousness is a particular kind of something it is like-ness that occurs only when certain conditions are met.
11. [^] In discussions of the nature and behavior of forces at the microscopic level we often find references to the way they ‘feel’ ([Feynman, 1963](#)), or the way they ‘experience’ each other in fields ([Rennie, 2015](#)). It would be interesting to investigate what motivates the use of such terms in this context.
12. [^] The difference between 1 and 0, for example, is an abstract difference conceived within a conscious mind.
13. [^] Arie Ben-Naim sets out in some detail how Shannon information is a probabilistic measure rather than a physical property ([Ben-Naim, 2015](#)). Note that the act of measurement presupposes a conscious mind capable of carrying out the measurement procedure and interpreting the result.

14. ^ Had he a fuller understanding of the nature of energy Bateson might not have been so dismissive about its role in mental processes. In *Mind and Nature* ([Bateson, 1979](#)) he referred only to kinetic energy (which he defined as “ MV^2 ”), thus ignoring potential energy, and was by his own admission “not up to date in modern physics.” In fact, slightly modifying Bateson’s much-cited phrase to *an actualized difference that makes a difference* yields a description of the essence of energetic action that is, the way energy, forces and work act antagonistically to effect change and cause further actions.
15. ^ Brains – as parts of people – process information in the colloquial sense, just as they process abstract ideas, equations, numbers, thoughts, emotions, or memories. But they do so as a *consequence* of the underlying energetic processing (conversion, distribution, and dissipation) going on in neural tissue. Computers also ‘process’ information in the colloquial sense. Mechanically and electronically speaking, however, they actually manipulate energy states (voltages, light, etc.) the results of which we, as conscious people, interpret informationally. It is worth noting that all mechanical information processing necessarily entails the dissipation of a certain amount of energy ([Landauer, 1961](#)). Recent experiments have confirmed this principle and demonstrated the intimate link between energy and what many refer to as information ([Bérut et al., 2012](#)).
16. ^ [Logan \(2012\)](#), in work undertaken with Stuart Kauffman and others, defines ‘biotic information’ as the organization of the exchange of energy and matter between organism and environment – a further example of information theory being used to quantify the biological organization of energy flows.
17. ^ It turns out this arrangement is energy efficient for the visual system overall ([Wong-Riley, 2010](#)).
18. ^ Examples can be found on YouTube.
19. ^ One way to quantify this relative complexity would be to follow the proposal of [Chaisson \(2001\)](#) and compare the energy rate density (a measure he calls Φ_m) between the two systems. Note also that Edelman and Gally are careful to distinguish cybernetic feedback in machine control systems from re-entrant processing in the brain, the latter being far more complex ([Edelman and Gally, 2013](#)).
20. ^ ‘I think therefore I am undergoing a certain recursive organization of actualized differences.’ Models of consciousness based on feedback loops in the brain have been discussed before, including by Douglas Hofstadter in his book *I am a Strange Loop* ([Hofstadter, 2007](#)). I have also previously proposed a feedback model of consciousness partly inspired by Edelman’s theory of reentrant processing as part of an attempt to design an artificially conscious work of art ([Pepperell, 2003](#)).
21. ^ The explanatory principle outlined here might be construed as a form of panpsychism or panexperientialism. My claim is not that consciousness is a fundamental property of nature, universally distributed. Rather, I claim it is a property of all physical systems that there is something it is like, intrinsically, to undergo actualized differences, a certain organization of which causes consciousness.

References

Abbott, M., and Van Ness, C. (1972). *Theory and Problems of Thermodynamics*, by *Schaum's Outline Series in Engineering*. New York, NY: McGraw-Hill Book Company.

[Google Scholar](#)

Aboy, M., Hornero, R., Abasolo, D., and Alvarez, D. (2006) Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans. Biomed. Eng.* 53, 2282–2288. doi: 10.1109/TBME.2006.883696

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Alkire, M. T., Hudetz, A. G., and Tononi, G. (2008). Consciousness and Anesthesia, *Science* 322, 876–880. doi: 10.1126/science.1149213

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Allen, J. (2009). *The Lives of the Brain: Human Evolution and the Organ of the Mind*. London: Harvard University Press. doi: 10.4159/9780674053496

[CrossRef Full Text](#) | [Google Scholar](#)

Annala, A. (2016). On the character of consciousness. *Front. Syst. Neurosci.* 10:27. doi: 10.3389/fnsys.2016.00027

[CrossRef Full Text](#) | [Google Scholar](#)

Aristotle (1818). *The Rhetoric, Poetic and Nicomachean Ethics of Aristotle* (tr. Thomas Taylor). London: James Black and Son.

[Google Scholar](#)

Aristotle (2002). *Metaphysics* (tr. Joe Sachs). Santa Fe, NM: Green Lion Press.

[Google Scholar](#)

Attwell, D., and Gibb, A. (2005). Neuroenergetics and the kinetic design of excitatory synapses. *Nat. Rev.* 6, 841–849. doi: 10.1038/nrn1784

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Bailey, D., Townsend, D., Valk, P., and Maisey, M. (2005). *Positron-Emission Tomography: Basic Sciences*. Secaucus, NJ: Springer-Verla. doi: 10.1007/b136169

[CrossRef Full Text](#) | [Google Scholar](#)

Bateson, G. (1972). *Steps Towards an Ecology of Mind*. San Francisco, CA: Chandler Publication Company.

[Google Scholar](#)

Bateson, G. (1979). *Mind and Nature: a Necessary Unity*. London: Wildwood.

[Google Scholar](#)

Bazzigaluppi, P., Amini, A. E., Weisspapier, I., Stefanovic, B., and Carlen, P. (2017). Hungry neurons: metabolic insights on seizure dynamics. *Int. J. Mol. Sci.* 18:E2269. doi: 10.3390/ijms18112269

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Ben-Naim, A. (2015). *Information, Entropy, Life and The Universe: What We Know and What We Do Not Know*. London: World Scientific. doi: 10.1142/9479

[CrossRef Full Text](#) | [Google Scholar](#)

Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R., and Lutz, E. (2012). Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* 483, 187–189. doi: 10.1038/nature10872.

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Bodart, O., Gosseries, O., Wannez, S., Thibaut, A., Annen, J., Boly, M., et al. (2017). Measures of metabolism and complexity in the brain of patients with disorders of consciousness. *Neuroimage Clin.* 14, 354–362. doi: 10.1016/j.nicl.2017.02.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Boltzmann, L. (1886). "The second law of thermodynamics," in *Ludwig Boltzmann: Theoretical Physics and Philosophical Problems: Selected Writings*, ed. B. McGinness (Dordrecht: D. Reidel).

[Google Scholar](#)

Buzsáki, G., Kaila, K., and Raichle, M. (2007). Inhibition and brain work, *Neuron* 56, 771–783. doi: 10.1016/j.neuron.2007.11.008

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Cahan, D. (ed.). (1995). *Hermann Von Helmholtz: Science and Culture*. Chicago: Chicago University Press.

[Google Scholar](#)

Capurro, R., and Hjørland, B. (2003). "The concept of information," in *Annual Review of Information Science and Technology (ARIST)*, ed. B. Cronin (Medford, NJ: Information Today), 343–411.

[Google Scholar](#)

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Trans. Med.* 5:198ra105. doi: 10.1126/scitranslmed.3006294

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Casarotto, S., Comanducci, A., Rosanova, M., and Sarasso, S (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* 80, 718–729. doi: 10.1002/ana.24779

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Chaisson, E. J. (2001). *Cosmic Evolution: The Rise of Complexity in Nature*. Cambridge, MA: Harvard University Press.

[Google Scholar](#)

Chatelle, C., Laureys, S., and Schnakers, C. (2011). “Disorders of consciousness: what do we know,” in *Characterizing Consciousness: From Cognition to the Clinic*, eds S. Deheane and Y. Christen (New York, NY: Springer).

[Google Scholar](#)

Chennu, S., Annen, J., Wannez, S., Thibaut, A., Chatelle, C., Cassol, H., et al. (2017). Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain* 140, 2120–2132. doi: 10.1093/brain/awx163

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Clarke, D. D., and Sokoloff, L. (1999). “Circulation and energy metabolism of the brain,” in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, eds B. Agranoff and G. Siegel (Philadelphia, PA: Lippincott-Raven), 637–670.

[Google Scholar](#)

Coelho, R. L. (2009). On the concept of energy: history and philosophy for science teaching. *Proc. Soc. Behav. Sci.* 1, 2648–2652. doi: 10.1016/j.sbspro.2009.01.468

[CrossRef Full Text](#) | [Google Scholar](#)

Collell, G., and Fauquet, J. (2015). Brain activity and cognition: a connection from thermodynamics and information theory. *Front. Psychol.* 6:818. doi: 10.3389/fpsyg.2015.00818.

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Coopersmith, J. (2010). *Energy, the Subtle Concept: The Discovery of Feynman's Blocks from Leibniz to Einstein*. Oxford: Oxford University Press.

[Google Scholar](#)

Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126. doi: 10.1038/nn0203-119

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Crutchfield, J. (1984). Space-time dynamics in video feedback. *Phys. D* 10, 229–245. doi: 10.1016/0167-2789(84)90264-1

[CrossRef Full Text](#) | [Google Scholar](#)

Davies, P. (2010). “Universe from bit,” in *Information and the Nature of Reality: From Physics to Metaphysics*, eds P. Davies and N. Gregersen (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511778759

[CrossRef Full Text](#) | [Google Scholar](#)

de Bono, M., and Maricq, A. V. (2005). Neuronal substrates of complex behaviors in *C. elegans*. *Annu. Rev. Neurosci.* 28, 451–501. doi: 10.1146/annurev.neuro.27.070203.144259

[CrossRef Full Text](#) | [Google Scholar](#)

Deacon, T. (2010). “What is missing from theories of information,” in *Information and the Nature of Reality: From Physics to Metaphysics*, eds P. Davies and N. Gregersen (Cambridge: Cambridge University Press).

[Google Scholar](#)

Deacon, T. (2013). *Incomplete Nature: How Mind Emerged From Matter*. New York, NY: W.W. Norton.

[Google Scholar](#)

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. London: Penguin Books.

[Google Scholar](#)

Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science* 358, 486–492. doi: 10.1126/science.aan8871

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Demertzi, A., Antonopoulos, G., Heine, L., Voss, H. U., Crone, J. S., Angeles, C. D. L., et al. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* 138, 2619–2631. doi: 10.1093/brain/awv169

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

de-Wit, L., Alexander, D., Ekroll, V., and Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychon. Bull. Rev.* 23, 1415–1428. doi: 10.3758/s13423-016-1002-0

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Dinuzzo, M., and Nedergaard, M. (2017). Brain energetics during the sleep–wake cycle. *Curr. Opin. Neurobiol.* 47, 65–72. doi: 10.1016/j.conb.2017.09.010

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Duncan, T. (2002). *Advanced Physics*. London: John Murray.

[Google Scholar](#)

Edelman, G. (2004). *Wider Than the Sky: The Phenomenal Gift of Consciousness*. New Haven, CT: Yale University Press.

[Google Scholar](#)

Edelman, G., Gally, J., and Baars, B. (2011). Biology of consciousness. *Front. Psychol.* 2:4. doi: 10.3389/fpsyg.2011.00004

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Edelman G., and Gally J. (2013). Reentry: a key mechanism for integration of brain function. *Front. Integr. Neurosci.* 7:63. doi: 10.3389/fnint.2013.00063

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Ellrod, F. E. (1982). Energeia and process in Aristotle. *Int. Phil. Q.* XXII, 175–181. doi: 10.5840/ipq198222216

[CrossRef Full Text](#) | [Google Scholar](#)

Erra, R., Mateos, D., Wennberg, R., and Perez Velazquez, J. (2016). Statistical mechanics of consciousness: maximization of information content of network is associated with conscious awareness. *Phys. Rev. E* 94:052402.

[PubMed Abstract](#) | [Google Scholar](#)

Fechner, G. T. (1905). *The Little Book of Life After Death* (tr. Mary C. Wadsworth). Boston: Little, Brown & Co.

[Google Scholar](#)

Feynman, R. (1963). *Conservation of Energy*, in *The Feynman Lectures on Physics*. Available at: http://www.feynmanlectures.caltech.edu/I_04.html [accessed December 21, 2017].

[Google Scholar](#)

Friston, K. (2013). Consciousness and hierarchical inference. *Neuropsychanalysis* 15, 38–42. doi: 10.1080/15294145.2013.10773716

[CrossRef Full Text](#) | [Google Scholar](#)

Gay, P. (1988). *Freud: A Life for Our Time*. London: W. W. Norton & Co.

[Google Scholar](#)

Goodale, M. (2014). How (and why) the visual control of action differs from visual perception. *Proc. Biol. Sci.* 281:20140337 doi: 10.1098/rspb.2014.0337

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hall, C. N., Klein-Flugge, M. C., Howarth, C., and Attwell, D. (2012). Oxidative phosphorylation, not glycolysis, powers presynaptic and postsynaptic mechanisms underlying brain information processing. *J. Neurosci.* 32, 8940–8951 doi: 10.1523/JNEUROSCI.0026-12.2012

[CrossRef Full Text](#) | [Google Scholar](#)

Hameroff, S., and Penrose, R. (2014). Consciousness in the universe: a review of the 'Orch OR' theory. *Phys. Life Rev.* 11, 39–78. doi: 10.1016/j.plrev.2013.08.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hawking, S. (1988). *A Brief History of Time*. London: Bantam Books.

[Google Scholar](#)

Heisenberg, W. (1958). *Physics and Philosophy: The Revolution in Modern Science*. New York, NY: Harper & Brothers Publishers.

[Google Scholar](#)

Hine, R. (ed.). (2015). *Oxford Dictionary of Biology*. Oxford: Oxford University Press doi: 10.1093/acref/9780198714378.001.0001

[CrossRef Full Text](#) | [Google Scholar](#)

Hofstadter, D. (2007). *I am a Strange Loop*. New York, NY: Basic Books.

[Google Scholar](#)

Hu, S., Stead, M., Liang, H., and Worrell, G.A. (2009). "Reference signal impact on EEG energy," *Advances in Neural Networks – ISNN 2009 Lecture Notes in Computer Science*, eds W. Yu, H. He, and N. Zhang (Berlin: Springer Science & Business Media), 605–611. doi: 10.1007/978-3-642-01513-7_66

[CrossRef Full Text](#) | [Google Scholar](#)

Hubel, D., and Livingstone, M. (1987). Segregation of form, color, and stereopsis in primate area 18. *J. Neurosci.* 7, 3378–3415. doi: 10.1523/JNEUROSCI.07-11-03378.1987

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hudetz, A. G. (2012). General Anesthesia and human brain connectivity, *Brain Connect.* 2, 291–302. doi: 10.1089/brain.2012.0107

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Hudetz, A. G., and Mashour, G. A. (2016). Disconnecting consciousness: is there an anesthetic end-point? *Anesth. Analg.* 123, 1228–1240. doi: 10.1213/ANE.0000000000001353

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

James, W. (1907). The energies of men. *Philos. Rev.* 16, 1–20. doi: 10.2307/2177575

[CrossRef Full Text](#) | [Google Scholar](#)

Kahn, C. H. (1989). *The Art and Thought of Heraclitus*. Cambridge: Cambridge University Press.

[Google Scholar](#)

King, J. -R., Sitt, J. D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., et al. (2013). Information sharing in the brain indexes consciousness in noncommunicative patients, *Curr. Biol.* 23, 1914–1919. doi: 10.1016/j.cub.2013.07.075

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Koch, C. (2004). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford: Oxford University Press.

[Google Scholar](#)

Koponen, L. M., Nieminen, J. O., and Ilmoniemi, R. J. (2015). Minimum-energy coils for transcranial magnetic stimulation: application to focal stimulation. *Brain Stimul.* 8, 124–134. doi: 10.1016/j.brs.2014.10.002

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Korzybski, A. (1933). *Science and Sanity. An Introduction to Non-Aristotelian Systems and General Semantics*. New York, NY: The International Non-Aristotelian Library Pub. Co, 747–761.

[Google Scholar](#)

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* 5, 183–191. doi: 10.1073/pnas.1219672110

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Laughlin, S. (2001). Energy as a constraint on the coding and processing of information. *Curr. Opin. Neurobiol.* 11, 475–480. doi: 10.1016/S0959-4388(00)00237-3

[CrossRef Full Text](#) | [Google Scholar](#)

Laureys, S., Lemaire, C., and Maquet, P. (1999). Cerebral metabolism during vegetative state and after recovery to consciousness. *J. Neurol. Neurosurg. Psychiatry* 67, 121–133. doi: 10.1136/jnnp.67.1.121

[CrossRef Full Text](#) | [Google Scholar](#)

Lee, U., Kim, S., Noh, G. -J., Choi, B. -M., Hwang, E., Mashour, G. A., et al. (2009). The directionality and functional organization of frontoparietal connectivity during consciousness and anesthesia in humans *Conscious. Cogn.* 18: 1069–1078. doi: 10.1016/j.concog.2009.04.004

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Lewis, L. D., Weiner, V. S., Mukamel, E. A., Donoghue, J. A., Eskandar, E. N., Madsen, J. R., et al. (2012). Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness. *Proc. Natl. Acad. Sci.* 109, E3377–E3386. doi: 10.1073/pnas.1210907109

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Logan, R. (2012). What is information?: why is it relativistic and what is its relationship to materiality. *Mean. Organ. Inform.* 3, 68–91.

[Google Scholar](#)

Lombardi, O., Hernán, F., and Vanni, L. (2016). What is shannon information? *Synthese* 193, 1983–2012. doi: 10.1007/s11229-015-0824-z

[CrossRef Full Text](#) | [Google Scholar](#)

Lotka, A. J. (1922). Contribution to the energetics of evolution. *Proc. Natl. Acad. Sci.* 8, 147–151. doi: 10.1073/pnas.8.6.147

[CrossRef Full Text](#) | [Google Scholar](#)

Magistretti, P. (2013). *Food for Thought: What Fuels Brain Cells? Report for Dana Alliance*. Available at: [http://www.dana.org/Publications/ReportOnProgress/Food for Thought What Fuels Brain Cells /](http://www.dana.org/Publications/ReportOnProgress/Food_for_Thought_What_Fuels_Brain_Cells/) [accessed March 27, 2018]

[Google Scholar](#)

Magistretti, P. J. (2008). “Brain energy metabolism,” in *Fundamental Neuroscience*, eds L. R. Squire, D. Berg, F. E. Bloom, S. du Lac, A. Ghosh, and N. C. Spitzer (Burlington, MA: Elsevier).

[Google Scholar](#)

Magistretti, P. J., and Allaman, I. (2013). “Brain energy metabolism,” in *Neuroscience in the 21st Century*, ed. D. W. Pfaff (New York, NY: Springer), 1591–1620. doi: 10.1007/978-1-4614-1997-6_56

[CrossRef Full Text](#) | [Google Scholar](#)

Marchetti, G. (2018) Consciousness: a unique way of processing information. *Cogn. Process.* 19, 435–464. doi: 10.1007/s10339-018-0855-8

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Mashour, G. A. (2004). Consciousness unbound. *Anesthesiology* 100, 428–433. doi: 10.1097/00000542-200402000-00035

[CrossRef Full Text](#) | [Google Scholar](#)

Morcom, A. M., and Fletcher, P. C. (2007). Does the brain have a baseline? Why we should be resisting a rest, *Neuroimage* 37, 1073–1082. doi: 10.1016/j.neuroimage.2006.09.013

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Morowitz, H. (1979). *Energy Flow in Biology*. Woodbridge, NJ: Ox Bow Press.

[Google Scholar](#)

Morowitz, H., and Smith, E. (2007). Energy flow and the organization of life. *Complexity* 13, 51–59. doi: 10.1002/cplx.20191

[CrossRef Full Text](#) | [Google Scholar](#)

Nagel, T. (1974). *What Is It Like to Be a Bat?* *Philos. Rev.* 83, 435–450.

[Google Scholar](#)

Niedermeyer, E., and Lopes da Silva, F. (1987). *Electroencephalography: Basic Principles, Clinical Applications, and Related Field*. Munich: Urban & Schwarzenberg.

[Google Scholar](#)

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Owen, A., Coleman, M., Boly, M., Davis, M., Laureys, S., and Pickard, J. (2006). Detecting awareness in the vegetative state. *Science* 313:1420. doi: 10.1126/science.1130197

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Pai, A., and Heining, M. (2007). Ketamine. *Contin. Educ. Anaesth. Crit. Care Pain* 7, 59–63. doi: 10.1093/bjaceaccp/mkm008

[CrossRef Full Text](#) | [Google Scholar](#)

Pepperell, R. (1995). *The Posthuman Condition: Consciousness Beyond the Brain*. Oxford: Intellect Books.

[Google Scholar](#)

Pepperell, R. (2003). Towards a conscious art. *Technoetic Arts* 1:2. doi: 10.1386/tear.1.2.117.18695

[CrossRef Full Text](#) | [Google Scholar](#)

Pepperell, R. (2018). “Art, energy and the brain,” in *The Arts and The Brain: Psychology and Physiology Beyond Pleasure (Progress in Brain Research)*, eds J. Christensen and A. Gomila (New York, NY: Elsevier).

[Google Scholar](#)

Perez Velazquez, J. (2009). Finding simplicity in complexity: general principles of biological and non- biological organization. *J. Biol. Phys.* 35, 209–221. doi: 10.1007/s10867-009-9146-z

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Raichle, M. (2011). “Intrinsic activity and consciousness,” in *Characterizing Consciousness: From Cognition to the Clinic?* eds S. Deheane and Y. Christen (New York, NY: Springer).

[Google Scholar](#)

Raichle, M. E. (2010). The brain’s dark energy. *Sci. Am.* 302, 44–49. doi: 10.1038/scientificamerican0310-44

[CrossRef Full Text](#) | [Google Scholar](#)

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Rennie, R. (ed.). (2015). *Oxford Dictionary of Physics*. Oxford: Oxford University Press.

[Google Scholar](#)

Rennie, R. (ed.). (2016). *Oxford Dictionary of Chemistry*. Oxford: Oxford University Press. doi: 10.1093/acref/9780198722823.001.0001

[CrossRef Full Text](#) | [Google Scholar](#)

Riehl, J. R., Palanca, B. J., and Ching, S. (2017). High-energy brain dynamics during anesthesia-induced unconsciousness. *Netw. Neurosci.* 1, 431–445. doi: 10.1162/NETN_a_00023

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Roederer, J. G. (2016). “Pragmatic information in biology and physics,” *Philos. Trans. A Math. Phys. Eng. Sci.* 374:20150152. doi: 10.1088/0034-4885/79/5/052601

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Rose, D. (1986). *Learning About Energy*. New York, NY: Plenum Press. doi: 10.1007/978-1-4757-5647-0

[CrossRef Full Text](#) | [Google Scholar](#)

Roy, C., and Sherrington, C. (1890). On the regulation of the blood supply in the brain. *J. Physiol.* 11, 85–108, 158–7–158-17. doi: 10.1113/jphysiol.1890.sp000321

[CrossRef Full Text](#) | [Google Scholar](#)

Ruffini, G. (2017). An algorithmic information theory of consciousness, *Neurosci. Conscious.* 3:nix019. doi: 10.1093/nc/nix019

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Schneider, E., and Sagan, D. (2005). *Into the Cool: Energy Flow, Thermodynamics, and Life*. Chicago, IL: University of Chicago Press.

[Google Scholar](#)

Schölvinck, M. L., Howarth, C., and Attwell, D. (2008). The cortical energy needed for conscious perception. *Neuroimage* 40, 1460–1468. doi: 10.1016/j.neuroimage.2008.01.032

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Schrödinger, E. (1944). *What is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press.

[Google Scholar](#)

Searle, J. (2013). *Can Information Theory Explain Consciousness*. Cambridge, MA: MIT Press.

[Google Scholar](#)

Seth, A. (2014). What behaviourism can (and cannot) tell us about brain imaging. *Trends Cogn. Sci.* 18, 5–6. doi: 10.1016/j.tics.2013.08.013

[CrossRef Full Text](#) | [Google Scholar](#)

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, doi: 10.1002/j.1538-7305.1948.tb01338.x

[CrossRef Full Text](#) | [Google Scholar](#)

Sherrington, C. (1940). *Man on his Nature*. London: Pelican Books.

[Google Scholar](#)

Shulman, G. L., Corbetta, M., Buckner, R. L., Fiez, J. A., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: I. Increases in subcortical structures and cerebellum but not in nonvisual cortex. *J. Cogn. Neurosci.* 9, 624–647. doi: 10.1162/jocn.1997.9.5.624

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Shulman, R., Hyder, F., and Rothman, D. (2009). Baseline brain energy supports the state of consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11096–11101. doi: 10.1073/pnas.0903941106

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Shulman, R. G. (2013). *Brain Imaging: What It Can (and Cannot) Tell Us About Consciousness*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199838721.001.0001

[CrossRef Full Text](#) | [Google Scholar](#)

Smil, V. (2008). *Energy in Nature and Society: General Energetics of Complex Systems*. Cambridge, MA: MIT Press.

[Google Scholar](#)

Stender, J., Mortensen, K. N., Thibaut, A., Darkner, S., Laureys, S., Gjedde, A., et al. (2016). The minimal energetic requirement of sustained awareness after brain injury. *Curr. Biol.* 26, 1494–1499. doi: 10.1016/j.cub.2016.04.024

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Sterling, P., and Laughlin, S. (2017). *Principles of Neural Design*. Cambridge, MA.: MIT Press.

[Google Scholar](#)

Strawson, G. (2008). “Real materialism,” in *Real Materialism and Other Essays* (Oxford: Oxford University Press), 19–52. doi: 10.1093/acprof:oso/9780199267422.003.0002

[CrossRef Full Text](#) | [Google Scholar](#)

Strawson, G. (2017). ‘Physicalist panpsychism’ in *The Blackwell Companion to Consciousness*, 2nd edn, eds S. Schneider and M. Velmans (New York, NY: Wiley-Blackwell), 374–390.

[Google Scholar](#)

Street, S. (2016). Neurobiology as information physics. *Front. Syst. Neurosci.* 10:90. doi: 10.3389/fnsys.2016.00090.

[CrossRef Full Text](#) | [Google Scholar](#)

Tagliazucchi, E., Behrens, M., and Laufs, H. (2013). Sleep neuroimaging and models of consciousness, *Front. Psychol.* 4:259. doi: 10.3389/fpsyg.2013.00256

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Arch. Ital. Biol.* 150, 290–326.

[Google Scholar](#)

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461. doi: 10.1038/nrn.2016.44

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Tozzi, A., Zare, M., and Benasich, A. (2016). New perspectives on spontaneous brain activity: dynamic networks and energy matter. *Front. Hum. Neurosci.* 10:247. doi: 10.3389/fnhum.2016.00247

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Vanhaudenhuyse, A., Noirhomme, Q., Tshibanda, L. J.-F., Bruno, M.-A., Boveroux, P., Schnakers, C., et al. (2009). Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain* 133, 161–171. doi: 10.1093/brain/awp313

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

von Foerster, H. (2003). “Notes on an epistemology for living things,” in *Understanding Understanding: Essays on Cybernetics and Cognition* (New York, NY: Springer).

[Google Scholar](#)

Werner, G. (2011). Letting the brain speak for itself. *Front. Physiol.* 2:60. doi: 10.3389/fphys.2011.00060

[CrossRef Full Text](#) | [Google Scholar](#)

Wiener, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.

[Google Scholar](#)

Wollstadt, P., Sellers, K. K., Rudelt, L., Priesemann, V., Hutt, A., Fröhlich, F., et al. (2017). Breakdown of local information processing may underlie isoflurane anesthesia effects. *PLoS Comput. Biol.* 13:e1005511. doi: 10.1371/journal.pcbi.1005511.

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Wong-Riley, M. (2010). Energy metabolism of the visual system. *Eye Brain* 2, 99–116. doi: 10.2147/EB.S9078

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Zhou, S., and Yu, Y. (2018). Synaptic E-I balance underlies efficient neural coding, *Front. Neurosci.* 12:46. doi: 10.3389/fnins.2018.00046

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

Ziv, J., and Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theory* 23, 337–343. doi: 10.1109/TIT.1977.1055714

[CrossRef Full Text](#) | [Google Scholar](#)

Keywords: consciousness, metabolism, energy, brain, information theory, feedback

Citation: Pepperell R (2018) Consciousness as a Physical Process Caused by the Organization of Energy in the Brain. *Front. Psychol.* 9:2091. doi: 10.3389/fpsyg.2018.02091

Received: 04 May 2018; **Accepted:** 10 October 2018;

Published: 01 November 2018.

Bridging the gap between consciousness and matter: recurrent out-of-body projection of visual awareness revealed by the law of non-identity

Published: 24 May 2023

Abstract

Consciousness is the most precious function of brain; however, there is an explanatory gap between consciousness and matter, which is deemed to affect the scientific research on consciousness. We believe that a methodological trap commonly present in scientific research and the incompleteness of logic are the true reasons that affect the research on consciousness. Here, a novel logic tool, the non-identity law, was extracted from physics and applied into the analysis of the visual dynamics related to naturalistic observation of night-shot still life, whose methodological approach is consistent with Descartes' matter-body-mind approach, breaking free from the methodological trap of current research. We show that visual system, the representative sensory system, has a postponed, recurrent out-of-body projection pathway from brain to observed object, besides the well-known feedforward signaling pathway available in existing literature, suggesting that human possesses an instinct of not only subjectively imaging (brain-generated imagery) but also projecting the image back onto the original or a particular place according to the clue of the manipulated afferent messenger light pathway. This finding adds a key piece of puzzle to the visual system. The out-of-body projection, coupled with neural correlates of consciousness (NCC), bridges the gap between consciousness and matter. This study in a self-contained and systematic manner provides a foundation for understanding the subjectivity and intentionality of human consciousness from the angle of visual awareness as well as the isomorphic relationships between unknowable original, private experience, and shareable expression (recording, calculus and deduction), showing that consciousness is obedient to certain rules rather than being unruly. The result paves the way for scientific research on consciousness and facilitates the integration of humanities and natural science.

Introduction

Consciousness, which is closely related to sensing, emotion, and thinking, is the focus of the mind-body problem long argued by philosophers from different philosophical schools, such as animism, dualism, materialism, idealism, and transcendental philosophy. To date, however, it has remained difficult to define consciousness clearly, and therefore, still have been the subject of intense debate within the philosophical community (Blackmore, [2006](#)). Contemporary

philosophers have argued that “consciousness is just about the last surviving mystery” (Dennett, [1991](#)) and that “consciousness is the biggest mystery” and “may be the largest outstanding obstacle in our quest for a scientific understanding of the universe” (Chalmers, [1996](#)).

Due to the close relationship between neuroscience and consciousness, neurobiologists were among the first to initiate scientific research on consciousness (Eccles, [1951](#); Edelman, [1989](#); Crick & Koch, [1990](#), [2003](#); Koch [2004](#)). There are four closely related questions that require an answer: 1) how does unconscious neural activity respond to an afferent sensory stimulus or memory recall, 2) where does consciousness arise, 3) how does consciousness arise, and 4) what is consciousness? Significant progress (such as in the cell and molecular biology of the neuron, synaptic transmission, and the neural basis of sensation, memory, cognition, perception, and movement) has been made regarding the first question over the past century (Albright et al., [2000](#); Kandel et al., [2012](#)). Regarding the second question, scientific research on the biological substrate of consciousness has identified the most likely neural correlates of consciousness (NCC) — the minimum neural mechanisms jointly sufficient for any one specific conscious experience (Koch et al., [2016](#)). There is, however, a debate as to whether the NCC is local or global (Mashour, [2018](#)). Although somewhat controversial at present, these two questions can be experimentally determined. Undoubtedly, the advances in neuroscience have greatly enriched our understanding of the neural activity concerning consciousness. However, they yet can not answer the last two questions mentioned above, termed as the “hard problem” of consciousness (Chalmers, [1995](#)).

This “hard problem” has drawn more scientists, including neurocomputing scientists and physicists, into this field. As a result, several theories of consciousness, such as the global workspace theory, integrated information theory, and Orch-OR theory have been proposed, each of which, in their own way, explain some aspects of consciousness (Van Gulick, [2018](#)). These theories, however, are neither derived from each other nor provide an adequate explanation of the nature of consciousness on their own. In turn, it is difficult to prove any of these theories unequivocally, thereby giving rise to some controversies (Michael, [2015](#); Koch & Hepp, [2006](#)). It remains difficult to give a scientific (quantitative) definition of consciousness and failing to reach a consensus.

All the attempts are the effects of the Monte Carlo method (Hammersley & Handscomb, [1965](#)) being applied in the academic community, which are necessary and important. However, existing forward and backward reasoning algorithms based on the semantics, including Monte Carlo method and Bayesian analysis (Bayes, [1763](#)), have not left room for unknown things, so it is impossible to infer unknown things from known facts in existing repository. Meanwhile, other research involving consciousness is relegated to a situation wherein the research is either conducted with only a vague notion of what consciousness means or by neglecting consciousness outright. For instance, frustration with interdisciplinary integration has raised concerns about the future of cognitive science enterprises (Núñez et al., [2019](#)). A fundamental error in cognitive science is that consideration of consciousness is neglected (Searle, [1990](#)). Without consciousness, there is no cognition. In a sense, overall, the crux of all scientific research on

consciousness points to the “hard problem” of consciousness. The problem comes full circle, and hence, must be faced directly.

Descartes asserted, “I doubt, therefore I think, therefore I am” (Descartes, [1637](#)). Since consciousness is a fact, why have previous attempts failed to explain consciousness? There must be unknown factors that limit the research on consciousness. We believe that there are two gaps rather than one gap in the study of consciousness, or, more precisely, one is a well-known gap and another is a methodological trap. First, consciousness is private and subjective, leading to the fathomless abyss, the great chasm or the explanatory gap between inner and outer, mind and brain, or subjective and objective (Blackmore, [2005](#), pp. 2; Levine, [1983](#)). Second, previous studies have suffered from a serious methodological limitation. Chalmers correctly pointed out that the conventional methods of cognitive science, neuroscience, and physicalism failed to explain the existence of consciousness (Chalmers, [1995](#), [1996](#)); however, he was unable to provide a feasible empirical research method. This is not a dilemma faced by a few people, but a dilemma faced by the whole field (Blackmore, [2006](#), pp. 1). We must find another way out.

We know that mathematics, a purely metaphysical system, is incomplete (Gentzen, [1938](#), [1969](#)). Is the empirical/positive study as to the real world also incomplete? We should first acknowledge a point of view that geometry is a physical science: assuming that a real thing named A has a boundary, there must be other real thing named B beyond the boundary of A, and it is the adjacent spaces of A and B that forms their boundaries (Clifford, [1886](#), pp. 47-49). In turn, without B, there is no boundary of A (i.e., A is deemed to be unlimited in the time-space dimension) and, hence, A is incompletely accessible to human observers. Thus, we can draw a conclusion that an empirical study that only concentrates on A itself is incomplete, called the incompleteness principle of empirical research (ICP) [Footnote1](#). The ICP demonstrates that there is a general methodological trap in empirical research. For the scientific research on consciousness, when we replace A with consciousness, by the ICP, the research that focuses only on consciousness itself is incomplete. Similarly, when regarded as a whole research object, the mind-body complex has an updated boundary. In this case, the research on mind-body complex confined to the complex itself is likewise incomplete. It is now clear that the methodological approach of most previous studies being confined in brain is far inferior to Descartes’ matter-body-mind approach (Descartes, [1637](#), pp. 99-108), implying previous studies have, at best, been able to identify related brain area and its neural spike to replace the pineal gland and animal spirits that were proposed by Descartes (Descartes, [1637](#), pp. 100).

Notice that the topic of the study, the first gap, is a Wittgenstein’s fact (Wittgenstein, [1921](#), pp. 5), it is not able to be changed. Alternatively, what we can do is to overcome the methodological trap. Then, how to eliminate the methodological drawback of ICP? We know that empirical research typically includes three steps: modeling, reasoning (deducing or disproving), and experimental verification. Obviously, if a study is considered incomplete, what we need to do is place the study within an expanded knowledge framework and reconstruct the model of empirical/positive study. Descartes’ matter-body-mind methodological approach has surpassed the mind-body approach and can meet this requirement without any modification.

Consequently, we plan to take four measures as following: 1) Acknowledging that consciousness is a fact, avoiding vague concepts such as soul, spirit, and free will — “what can be said at all can be said clearly, and what we cannot talk about we must pass over in silence” (Wittgenstein, [1921](#), pp. 3). 2) Attention is paid to philosophy, such as Socrates’, Lao-Tzu’s, Descartes’, Kant’s, and Wittgenstein’s philosophy, whose senses is doubting everything, even the authors themselves. Likewise, we should pay attention to the natural sciences from physics to neuroscience/cognitive neuroscience, absorb their insights about the world, and ensure that the perspectives in the article are compatible with them. 3) Reconstructing the model of empirical/positive study of consciousness according to Descartes’ matter-body-mind methodological approach. We noticed that almost all knowledge in natural and humanities sciences has been considered in previous research, with the only exception being logic, suggesting that we may need to do something in the field of logic.

Specifically, this study endeavored to explore what consciousness is from the perspective of factual visual awareness, in a systematic manner. First, a novel logic tool, called the non-identity law, was proposed based on physics to replace the law of identity. Second, the law of non-identity was applied to the analysis of expanded visual dynamics, building on a complete visual stimulation–response pathway related to the naturalistic observation of a night-shot still life, whose methodological approach is consistent with Descartes’ matter-body-mind approach. Finally, whether an out-of-body projection (OBP) is a physical behavior and the relations between an original, experience, and expression were discussed carefully. The results revealed that visual awareness, the representative sensation, can be understood in physical terms. Note that the term “visual awareness” in this paper is used in its broadest sense to refer to the entire process of sight, including various objective and subjective facts that occur along a complete visual pathway.

Theoretical preparation: the law of non-identity

Sensation, emotion, imaging, thinking, and movement are all facts. In contrast, logic is an abstract/metaphysical expression about the facts, written in symbols and words. It is a cognitive tool for understanding the relationships among things in the universe, in which the law of identity, one of the three basic laws, is denoted as

$$Q=Q, \text{ or } Q \rightarrow Q,$$

where Q is a thought object, such as a name, concept, event, or relation. The definition of the law is a tautology that states nothing at all about facts (Wittgenstein, [1921](#), pp. 63). Fundamentally, the law of identity is incompatible with the ICP introduced above.

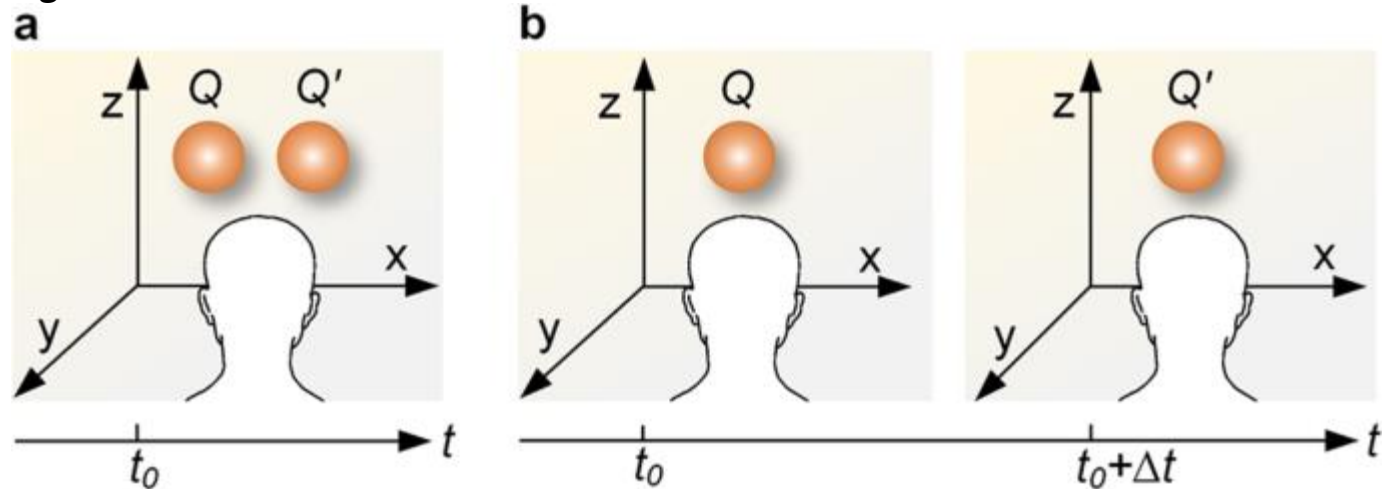
Here, we endeavor to challenge the law of identity within the frame of reference of natural science. Time, space, and matter are the three measurable interdependent elements that constitute the universe, while any object reveals itself (endurance, size, and position) by its interactions with other matter (Ridley, [1995](#), pp. 40–41, 85–86). Thus, an object can be denoted as

$$Q=Q(m,s,t),$$

where Q is the name of the object. Q depends on three measurable physical parameters, s , t , and m , in which s denotes the shape and location of Q , m the rest mass or relativistic mass (or the sum total of mass and energy) of Q , and t the time. These values are determined by the interaction of Q with its surroundings. For simplicity, unless otherwise stated, Q refers to $Q(m, s, t)$ throughout the paper, and the following names, Q' and Q'' , are referred to similarly.

First, consider any two self-governing condensed objects, $Q(m, s, t)$ and $Q'(m', s', t)$, that are at rest and located in different places in an inertial frame of reference (Fig. 1a). Macroscopic condensed matter consists of a large number of particles under four interaction forces. Suppose Q contains $J(t)$ elementary particles $\{q_1, q_2, \dots, q_{J(t)}\}$ at time t , where the i th particle q_i possesses the mass $m(i)$ and the occupying space $s(i)$. Thus, Q possesses the shape and location expressed by $s = \bigcup_{i=1}^{J(t)} s(i)$ and a mass denoted by $m = \sum_{i=1}^{J(t)} m(i)$. Similarly, Q' containing $K(t)$ elementary particles $\{q_1', q_2', \dots, q_{K(t)}'\}$ at time t possesses the shape and location expressed by $s' = \bigcup_{i=1}^{K(t)} s'(i)$ and a mass denoted by $m' = \sum_{i=1}^{K(t)} m'(i)$. This is the basic physical representation of the two objects at the atomic level. Let us now return to the beginning of this paragraph and consider the problem: why is the number of objects in the inertial frame of reference deemed to be two?

Fig. 1



Identifying an object in the inertial frame of reference. **a**, Discriminating two objects in the space dimension. An observer simultaneously observes two objects Q and Q' that are located at different places at the same time t_0 . **b**, Identifying one object in the time dimension. An observer sees an object Q at t_0 for the first time; after closing his/her eyes for Δt , the observer sees an object Q' at $t_0 + \Delta t$, where Q' is the continuation of Q in the time dimension.

Obviously, the spaces occupied separately by two macroscopic objects, whose borders and contours can be perceived by humans, have never overlapped with each other. In other words, the two macroscopic objects have no chance to meet or collide to exchange any substance with each other at any time; therefore, the two macroscopic objects do not share any particle at any time. Thus, the term “absolute non-identity law” can be defined as follows: two objects are called two objects (i.e., $Q \cap Q' = \emptyset$) if, and only if,

$$\forall(t\xi)(s(t\xi)\cap s'(t\xi)=\phi),$$

where Q and Q' are the names of the two objects, respectively, s and s' are the spaces occupied separately by Q and Q' , and t_ξ is the time variable that traverses throughout the observation time, beginning with the moment t_0 and ending with t_1 , i.e., $t_\xi \in [t_0, t_1]$. The absolute non-identity law shows how to distinguish one object from another. Note that although extracted from the inertial frame of reference with any two still objects, the absolute non-identity law can be generalized to the inertial frame of reference with any two objects in relative motion as long as their spaces are independent of each other at any one time.

Second, suppose that there is only one object in the inertial frame of reference (Fig. 1b); the observer observes an object $Q(m, s, t)$ at t_0 . Closing their eyes for Δt , the observer opens their eyes again and observes an object $Q'(m', s', t)$ at $t_1 = t_0 + \Delta t$ ($\Delta t \neq 0$). Is Q' identical to Q ?

Each spinning particle in Q interacts with every other particle, resulting in the creation, decay, and annihilation of the particle whose position and momentum cannot be, even in principle, determined precisely (Ridley, 1995, pp. 112–113). Furthermore, substances have been continuously adsorbed onto Q or have escaped from Q in the open system. Considering that both the spatial structure and mass of Q have been altered perpetually, the number of different particles in Q' observed at t_1 is not identical to that of Q at t_0 . Thus, the term “relative non-identity law,” can be defined as

$$\forall(t_0)\forall(t_1)(t_1>t_0)\rightarrow(Q'\cap Q\neq Q)\wedge(Q'\cap Q\neq Q')\wedge(Q'\cap Q\neq\emptyset),$$

where Q' (i.e., $Q'(t_1)$) is the continuation of Q (i.e., $Q(t_0)$) in the time dimension.

The absolute non-identity law provides a method for discriminating and naming one macroscopic object as distinguished from others in the perceivable space dimension^{Footnote2}, whereas the relative non-identity law indicates the evolution of one macroscopic object in the perceivable time dimension. Both laws, associated with the physical world, contain a greater reductionism connotation than the law of identity (in a sense, the law of non-identity is another expression reflecting the same aim as the law of identity). For example, by the relative non-identity law, we can readily solve some classic paradoxes, such as the ship of Theseus and the dispute about whether a man can step in the “same” river twice. However, since the absolute non-identity law, an unambiguous binary computing tool, is more rigorous than the relative non-identity law, we applied the former to the analysis of visual dynamics. In the following sections, we show that the non-identity law is essential as a reductionist tool for elucidating visual awareness.

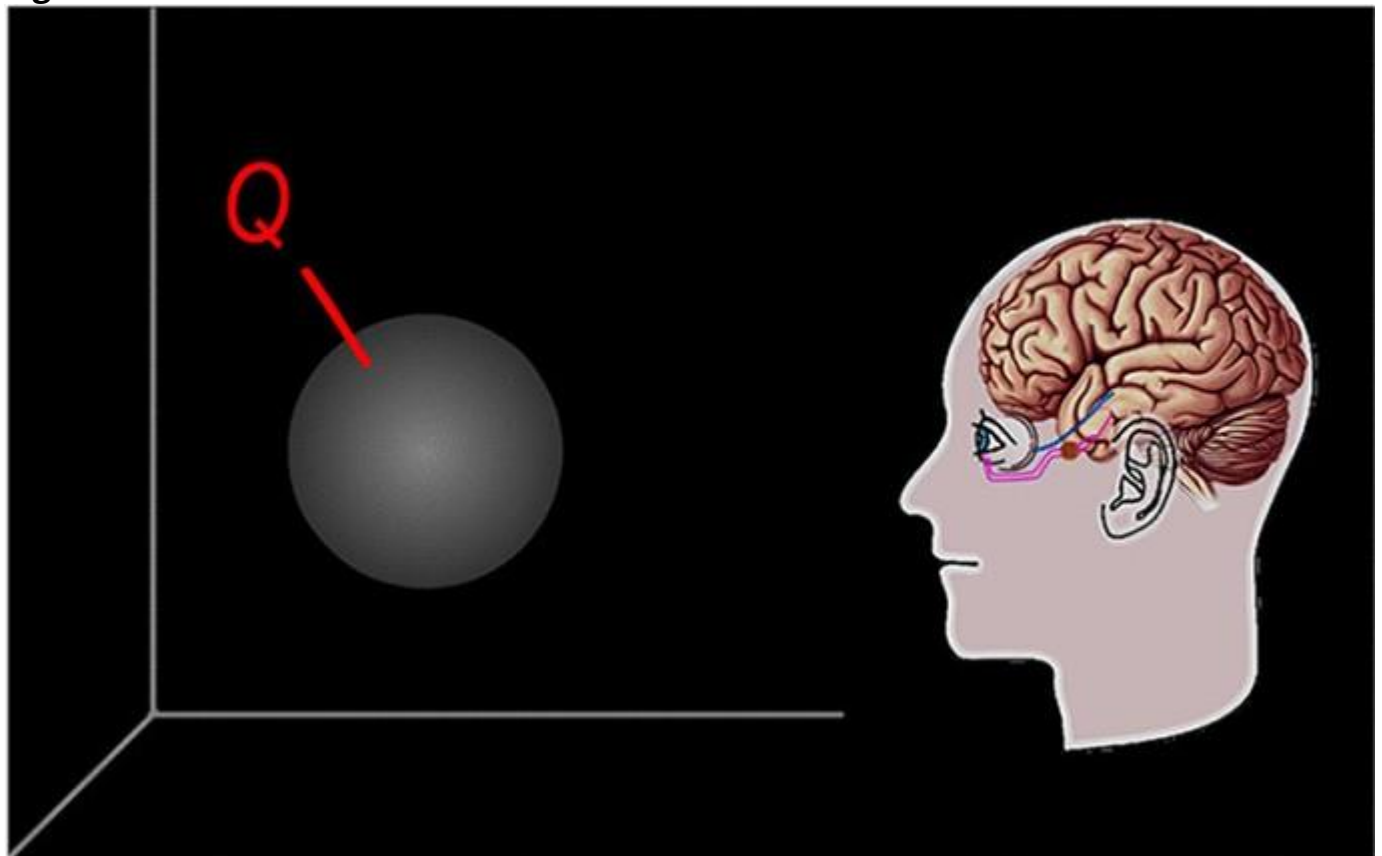
Classic projection of brain-generated imagery from brain to observed object

Global visual dynamics model related to naturalistic observation

Visual awareness is a vital component of consciousness, and its biological substrate involves almost half of the cerebral cortex. It has several advantages over other components for investigating the neural basis of consciousness (Crick & Koch, [1990](#)). In particular, it is scarcely affected by emotion. Significant progress has been made in elucidating visual dynamics in the brain (Walsh & Cowey, [1998](#); Lamme & Roelfsema, [2000](#); Tapia & Beck, [2014](#)). However, the contribution of the messenger of the observed object and the outcome of the visual system have been neglected. Therefore, an analysis of the visual dynamics described with first- and third-person data (Chalmers, [2013](#)) that arises over an expanded visual pathway, beginning with an observed object and ending with the visual outcome, is required. As this expanded visual pathway is global, it is also referred to as the “global visual pathway”; accordingly, its dynamics are called “global visual dynamics.”

Let us first build the global visual dynamics model. Suppose there is a ball in a dark room (Fig. [2](#)). The ball is denoted by $Q(m, s, t)$, whose definition is explained in the theoretical preparation section. At this moment, the ball is unknown to all observers; hence, it is referred to as the original or thing-in-itself. How does an observer perceive the original?

Fig. 2



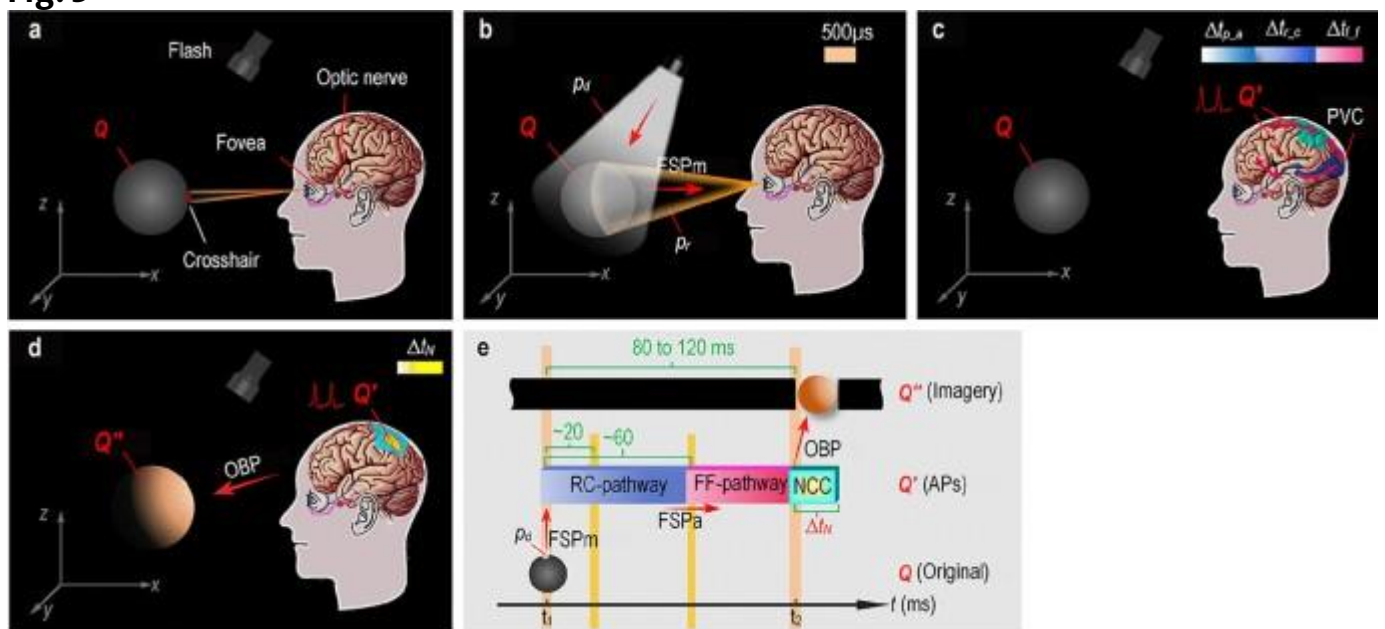
Object in the dark, unknown to any observer

It is well known that to explore an unknown thing actively, a stimulus–response test is usually required. Here, a specific research paradigm for the visual stimulus–response test was conceived,

in which the single-pulse diffuse reflection is a dark-to-light-to-dark stimulus, while the human visual system acts as a detector. Different from self-luminous masking, not only is this paradigm closer to the naturalistic observation, but it also has an added advantage—the darkness does not enable the retina to encode signals effectively, thereby ensuring no interference before and after the visual stimulus. In compliance with the paradigm, two visual psychophysical experiments of perceiving night-shot still life under the pre-induced accommodation condition were conducted. Eighteen volunteers from the University of Electronic Science and Technology of China were recruited for the experiments (14 males and four females who were aged between 18 and 49 years, were not colorblind, and had no known neurological or visual disorders; six of them had normal vision and the others had corrected-to-normal vision. The study was reviewed and approved by the Ethics Committee of the University of Electronic Science and Technology of China. The participants provided their written informed consent to participate in this study before inclusion in the experiment). The experimental result shows that each participant correctly reported the position, shape, and color of a constant or randomly selected still life after a single-pulse diffuse reflection stimulus lasting for 500 μ s under the pre-induced accommodation condition associated with visual attention (see Supplementary Materials and Methods, Supplementary Text, Supplementary Figures 1–4, and Supplementary Table 1 for details).

Having established a rough outline of the visual stimulus–response effect, let us now consider how the effect occurs in the detector (visual system). Based on the above visual psychophysical experiments, in conjunction with the neural dynamics in the extant literature, a complete visual dynamics model is presented (Fig. 3). In this model, the flash, when turned on at t_i , begins to emit broad-spectrum photons that strike the surface of Q , where the photons are called detective photons. Some of the detective

Fig. 3



Global visual dynamics in response to an afferent single-pulse messenger light signal. a, An accommodation pattern is built in advance using a weak LED (light-emitting diode) crosshair,

where the original Q and human perception are all in the dark. **b**, Messenger-light feedforward signaling pathway (FSPm). A flash has shined on Q for 500 us, when some detective photons (p_d) reflected by Q , p_r enter the eyes. **c**, AP feedforward signaling pathway (FSPa). After the flash onset, p_r converging at the retina have been encoded as action potentials (APs, i.e., Q') for Δt_{p_a} . Q' is submitted to the optic nerve, when the perception is still dark. Next, Q' has been relayed over the retinocortical pathway (RC-pathway) beginning at the retina and ending at the primary visual cortex in Δt_{r_c} . Then, it takes Δt_{f_f} for Q' to propagate sweepingly in a feedforward–feedback pathway (FF-pathway, red bidirectional long-dash ring). Finally, Q' feeds back into the posterior parietal cortex (PPC, Cyan area), while everything is in the dark. **d**, Out-of-body projection (OBP). After previous unconscious brain activity, a bright image Q'' arising from the PPC hot zone (the likely neural correlates of consciousness [NCC], yellow area) appears in front of the observer while the real world is still dark, and hence there is a recurrent OBP linking Q' to Q'' via the NCC. Subsequently, everything falls into darkness. **e**, Timing sequence of events related to vision. The reflected p_d shaping FSPm at t_1 has been encoded as Q' in the retina for about 20 ms (Kandel et al. 2012, pp. 594). Subsequently, Q' has propagated through the rest of FSPa to the NCC. Eventually, the perception Q'' of Q arises from the NCC at t_2 . The duration of the unconscious response, $t_2 - t_1$, is between 80 ms and 120 ms, in which the retinocortical transmission time Δt_{r_c} is around 60 ms (Walsh & Cowey, 1998; Tapia & Beck, 2014).

photons are absorbed and the rest are reflected around, some of which enter the eyes (Fig. 3b). The process can be described as follows:

$$\rightarrow pd(t_1)Q(m,s,t) \rightarrow pr(t_1),$$

where p_d represents the detective photons, and p_r is the stray light cone entering the eyes, carrying the absorption and reflection information of p_d interacting with Q at t_1 , i.e., the messenger light of Q . Given the complexity of particle physics, the space and mass parameters of the light were omitted for simplicity.

Subsequently, the retina performs light-to-electricity transduction: the messenger light induces the rods and cones to trigger chemical signals continuously. These signals are sent through bipolar cells to the retinal ganglion cells where the chemical signals are eventually encoded as trains of action potentials (APs) that facilitate long-distance information propagation via the optic nerve (Kandel et al., 2012; Bear et al., 2015). Some ganglion cells show a response peak at only 20 ms after the flash onset (Kandel et al., 2012, pp. 594). The light-to-electricity transformation that takes place in the retina can be described as follows:

$$\rightarrow pr(t_1)retina \rightarrow Q'(e',s',t_1+\Delta t_{p_a}),$$

where Q' denotes the encoding APs, e' the bioelectrical energy, s' the finite-dimensional traces of Q' , and Δt_{p_a} the light-to-electricity transformation latency of about 20 ms ^{Footnote3}.

Correspondingly, there is a messenger-light signaling pathway $Q \rightarrow prretina$, an external feedforward signaling pathway (FSP), which is termed FSPm.

Up to this point, the retina has accomplished low-level visual processing tasks, such as light-to-electricity transduction, acquisition, and iconic memory. At this point, image acquisition devices can directly project an image onto a screen for us. In comparison, where and how does the human visual system present what one sees for oneself?

First, the encoding Q' arising from the retina is conveyed through the lateral geniculate nucleus to the primary visual cortex (Fig. 3c), whose neural pathway is often referred to as retinofugal projection (Bear et al., 2015). Electrophysiological research has demonstrated that the retinocortical transmission time Δt_{r_c} varies from 55 ms to 70 ms and averages approximately 60 ms (Tapia & Beck, 2014).

Next, from the primary visual cortex, the encoded information is transmitted via two major pathways, first described in a monkey: a dorsal pathway into the frontal lobe through a number of extrastriate areas in the parietal lobe, and a ventral pathway into the frontal lobe through the inferior temporal cortex (Mishkin et al., 1983). Generally, the dorsal pathway processes the spatial information (position, motion, speed) related to fast visuo-motor control, whereas the ventral pathway processes information about the form (color, shape, texture) related to perception, both of which also exist in other sensory systems (Kandel et al., 2012). Additionally, the propagation of APs through lateral and feedback connections, in which the same cortical neurons contribute to different analyses at different moments in time, can be incorporated into a sweeping feedforward–feedback response (Lamme & Roelfsema, 2000).

Note that although unconscious, the above stage where Q' propagates sweepingly over the intracorporal FSP with conjunction and disjunction, is not only correlated with consciousness, but is also necessary for it (Tapia & Beck, 2014). Compared to image acquisition devices, the added sweeping feedforward–feedback response demonstrates that the brain makes an extraordinary effort to generate imagery.

At length, the distributed encodings linked to the different visual features (e.g., position, shape, and color) are eventually bound together (Crick & Koch, 1990, 2003; Treisman, 1996). As a result, a perception of Q with its various features arises from the NCC. The NCC is believed to be primarily localized to a posterior cortical hot zone that includes sensory areas (Koch et al., 2016); however, the distribution of the NCC remains controversial (Mashour, 2018). Considering that no consequence is altered by variations in the layout of the NCC, the statement of the posterior cortical hot zone was used in this study.

Thus far, the visual signaling pathway, $pd \rightarrow Q \rightarrow pr_{retina} \rightarrow Q'NCC$, has been completely presented. In contrast to the extraneous messenger-light pathway, the AP pathway $retina \rightarrow Q'NCC$, the intracorporal part of the FSP is termed FSPa. Previous research investigating whether perception is blocked by transcranial magnetic stimulation pulses applied to the early visual cortex demonstrates that the duration of unconscious brain activity after stimulus onset, Δt_{FSPa} (i.e., $\Delta t_{r_c} + \Delta t_{f_f}$), typically ranges from 80 ms to 120 ms (Walsh &

Cowey, 1998; Tapia & Beck, 2014), where Δt_{f-f} is the duration of the feedforward–feedback response. This observation is consistent with findings from other visual psychophysical studies. For instance, the critical flicker fusion frequency of humans is 60 Hz (Healy et al., 2013), while perception occurs approximately 80 ms after stimulus onset in the flash-lag effect experiment (Eagleman & Sejnowski, 2000). In other words, these experimental methods for measuring the cycle of visual dynamics are equivalent.

The visual experience of an afterimage can persist for several minutes, helping us to infer the duration of the NCC activity without examining a biomarker of NCC activity. Contrastingly, the duration of visual perception of a night-shot still life image is very short, and therefore the duration of NCC activity, Δt_N , should be accurately determined by checking the NCC activity. However, the NCC remains to be finalized; hence, Δt_N remains unknown. Instead, another issue of greater concern is the output of physical visual processing.

It should be considered that the migrating APs carrying all visual information on Q are situated inside the brain (Fig. 3c); hence, the output of visual processing should likewise be confined to the brain. However, interestingly, although the real world in this experiment is dark, an orange ball Q'' appears in front of the observer, rather than inside the brain of the observer (Supplementary Figures 3–4, and Fig. 3d). Although short-lived, Q'' is a real visual experience for the observer. No matter what it is, Q'' can always be expressed as

$$Q''(m'', s'', c, t_2),$$

where Q'' denotes what the observer sees in front of him/her; it can be described by four parameters: m'' , the unknown mass of Q'' ; s'' , the space occupied by Q'' ; c , the color of Q'' ; and t_2 (i.e., $t_1 + \Delta t_{FSPa}$), the time at which the perception emerges from the NCC (Fig. 3e).

There is no doubt that a transformation (also called emergence) has occurred in the NCC:

$$\rightarrow Q'(e', s', t_2)NCC \Rightarrow Q''(m'', s'', c, t_2).$$

Analogous to the projectors, the behavioral manifestation, $NCC \Rightarrow Q''$, can be termed an OBP. So far, we have mainly presented the global visual dynamics, based on which we can conclude that the visual system possesses an FSP-OBP pathway,

$$pd \rightarrow Q \rightarrow prretina \rightarrow Q'NCC \Rightarrow Q'',$$

where $pd \rightarrow Q \rightarrow prretina \rightarrow Q'NCC$ is the unconscious FSP, $NCC \Rightarrow Q''$ is the OBP pathway of the conscious outcome Q'' , the brain-generated imagery in response to the afferent messenger light of a particular original.

Visual dynamics analysis with non-identity law

Let us now analyze the relationships between Q , Q' , and Q'' . First, the encoding Q' of the messenger light and the original Q are located at both ends of the FSPm pathway, i.e., Q at rest is located at an extracorporeal place, while Q' , in response to some detective photons reflected by Q , is located inside the body. Hence,

$$\forall(t\xi)(s(t\xi) \cap s'(t\xi) = \emptyset),$$

where s and s' are the spaces owned separately by Q and Q' , and t_ξ is the time variable that traverses throughout the observation period. By the non-identity law described in the theoretical preparation section, we obtain

$$Q \cap Q' = \emptyset,$$

i.e., Q' and Q are two different things.

Similarly,

$$Q'' \cap Q' = \emptyset.$$

Q'' and Q' are, therefore, two different things. Then, what is Q'' ? This could simply be answered by announcing a new concept. However, after searching the existing literature, we are convinced that an OBP, the conscious outcome of the visual system, i.e., the brain-generated imagery, is the so-called visual awareness in a narrow sense. In other words, what we see in front of us is precisely our visual awareness. Notably, although we know the timing sequences of Q , Q' , and Q'' , the spatial connection between Q and Q' , and the spatial connection between Q' and Q'' , we are unable to deduce the spatial connection between Q'' and Q directly by the equations $Q \cap Q' = \emptyset$ and $Q'' \cap Q' = \emptyset$. Consequently, the next step is to confirm their spatial connection quantitatively.

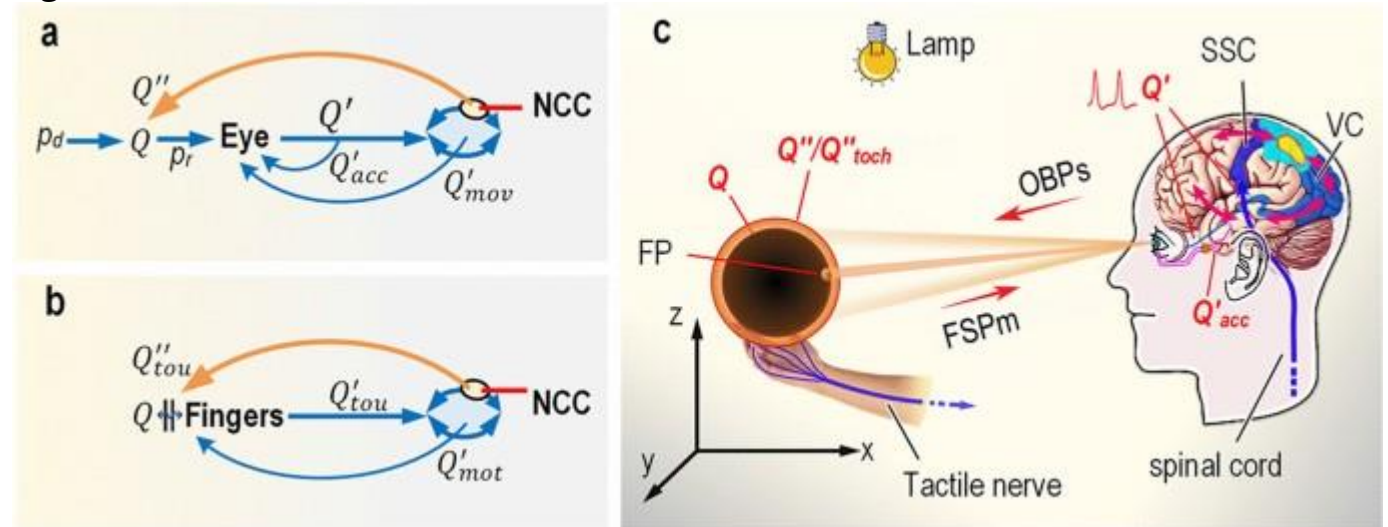
Projection position reciprocally determined by vision and touch

As described previously, the dorsal pathway determines where the external thing is, whereas the ventral pathway determines what it is. Visuospatial information regarding gaze direction and gaze distance is common to all visual cortical areas (Dobbins et al., [1998](#)). The messenger light of an original, however, cannot carry visuospatial information other than the absorption and reflection information of detective photons interacting with an original. Thus, it is natural to question where the visuospatial information originates.

The visual system possesses a mechanism to generate the visuospatial information by itself, involving a set of sophisticated dynamic optical control mechanisms comprising eye movement, pupillary reflex, and accommodation (Kandel et al. [2012](#), pp. 562). In principle, a dynamic system can be regarded as an automatic regulatory system. When the mechanism is introduced into the global visual dynamics model, a symbolized semi-closed-loop visual dynamics model,

the automatic regulatory system with two typical transformations that occur separately at the eye and NCC can be presented (Fig. 4a). In this symbolized model, an optical focus is achieved by applying the top-down encoding Q_{mov}' for eye movement and the local encoding Q_{acc}' for accommodation to the extraocular and ocular muscles, where the encodings are correlated with the direction and distance of the observed object Q . In turn, the visuospatial information on the direction and distance of Q can be represented by both encodings. Thus, it can be inferred that the sites at which Q_{mov}' and Q_{acc}' are encoded for an optical focus provide a copy of the encodings to all visual cortical areas for intermediate- and high-level visuospatial processing, in compliance with size constancy in terms of shape, motion, and speed (Kandel et al., 2012; Schwartz et al., 1983).

Fig. 4



Visual-tactile intercalibration based on visual and tactile automatic regulatory systems.

a, Semi-closed-loop visual regulatory system. When an observer stares at Q , the encoding Q' of the messenger light of Q , which arises from the retina, is sweepingly propagated to the feedforward–feedback circuitry (ellipsoid). Consequently, a top-down encoding Q_{mov}' for eye movement control innervates the extraocular muscles to move the eyeball, pointing the fovea towards the fixation point (FP) on Q , whereas the local encoding Q_{acc}' for accommodation, which derives from the optic nerve, innervates the ocular muscles to regulate the pupil and lens. As a result, the controlled optical system only allows a narrow beam of messenger light of the FP to focus on the fovea and collaterally allows the messenger light of Q to focus on the retina. At length, a sharp image Q'' is projected back to Q by the neural correlates of consciousness (NCC). **b**, Closed-loop tactile regulatory system. The encoding Q_{mot}' coming from the motor area controls the coordinated arm–wrist–finger motion until the fingers touch an object. Consequently, the encoding Q_{tou}' of the electromagnetic interaction occurring at the interface $Q \parallel$ Fingers is fed back through the feedforward–feedback circuitry into the NCC, by which a tactile experience Q_{tou}'' is projected to the interface. **c**, Visual–tactile intercalibration. The reflected photons shaping the messenger-light feedforward signaling pathway (FSPm) reach the retinae, at which Q' is encoded and conveyed to the visual cortex (VC). The tactile signal encoding, through the tactile nerve and spinal cord, enters the somatosensory cortex (SSC). Both signals propagate sweepingly through the feedforward–feedback pathway to the NCC. At length, an image Q'' is projected onto Q and is confirmed by the tactile perception Q_{tou}'' .

In fact, spatial cognition is achieved by visuotactile integration training (Batista et al., 1999; Gentile et al., 2011; Chen et al., 2016). When touching an object, one can shape a tactile experience $Q_{tou}''(p_{tou}'', s_{tou}'', k_{tou}'', t)$ of Q , where p_{tou}'' is the perceptual pressure in response to the electromagnetic interaction that occurs at the contact interface and k_{tou}'' is the perceptual temperature of Q . Although many more objective and subjective parameters can be adopted, only four parameters are used here for simplicity. Obviously, the tactile outcome is not located inside the brain, but at the contact interface. Therefore, similar to the symbolized visual dynamics model, we can obtain a symbolized closed-loop tactile dynamics model: the automatic regulatory system with two typical transformations that occur separately at the NCC and the finger receptors (Fig. 4b).

In contrast to the semi-closed-loop visual system, tactile receptors, the farthest node of the closed-loop pathway of touch, can directly approach an object through the moveable arm and fingers (Fig. 4c). For example, to explore the outside world, the top-down encoding Q_{mot}' , encoded by the motor system, voluntarily drives the arm and fingers until the fingers touch an object. Hence, there is a contact interface $Q \parallel \text{Finger}$ at which the electromagnetic interaction occurs. Subsequently, the encoding Q_{tou}' of electromagnetic interaction feeds back into the brain and mediates the NCC to project a tactile perception Q_{tou}'' to the contact interface. Thus, Q_{tou}'' spatially accords with Q , i.e., $s_{tou}'' \approx s$. Moreover, there is evidence that spatial information of the same object obtained by sight and touch can be calibrated with each other (Chen et al., 2016), implying that $s'' \approx s_{tou}''$. Therefore, $s'' \approx s$. This demonstrates that the visual OBP is spatially superimposed onto the observed object, creating a complete visual causal chain as shown in Fig. 4a.

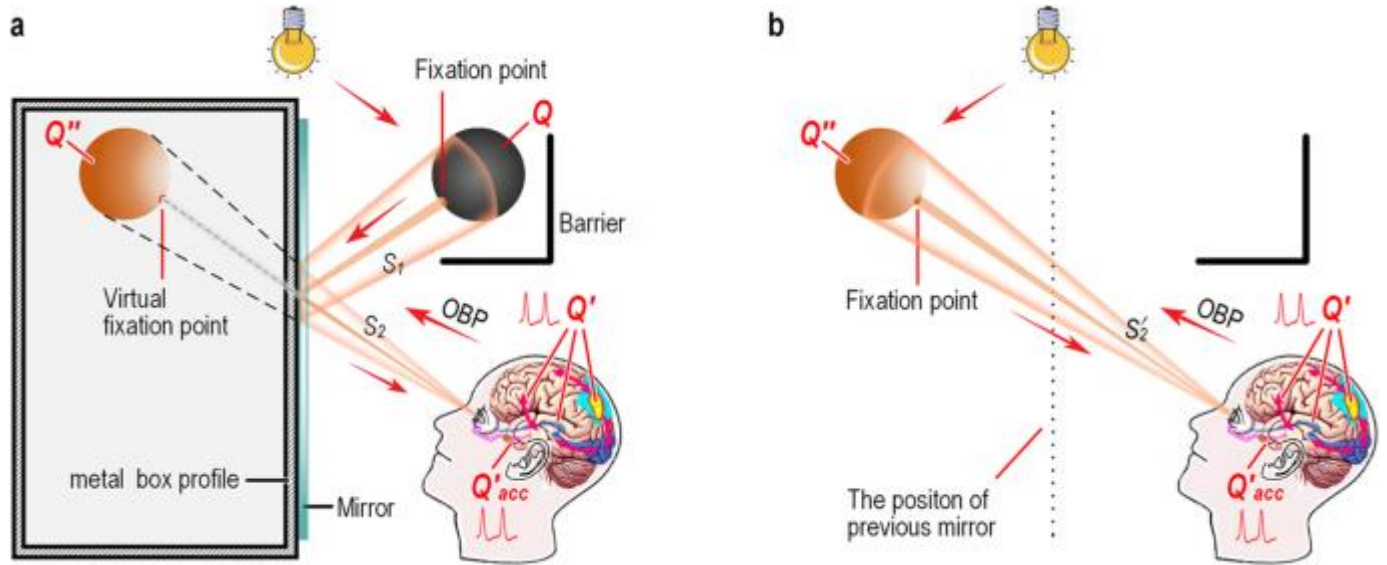
Nontrivial projection of brain-generated imagery separated from its original

As described previously, a classic visual pathway, $pd \rightarrow Q \rightarrow prretina \rightarrow Q'NCC \Rightarrow Q''$, where Q'' is usually superimposed onto its original Q can be reciprocally verified by visual and tactile perception, facilitating the belief that what one sees is exactly the object itself. However, another instance that we typically encounter on a daily basis should challenge this impression: the mirror image.

To revisit the mirror image, let us conceive a thought experiment on the mirror image (Fig. 5a). When an observer looks into a mirror, the messenger-light pathway

of Q is $pd \rightarrow Q \rightarrow prmirror \rightarrow prretina$, where the FSPm is deflected by the mirror. Hence, there are two segments of the messenger-light pathway: S_1 , a pathway from Q to the mirror, and S_2 from the mirror to the retina. The messenger light converging on the retina is encoded as Q' , which enters the brain through the optic nerve for intermediate- and high-level visual processing. To present the image Q'' behind the mirror clearly, the top-down encoding Q_{mov}' innervates the extraocular muscles to point the fovea to S_2 , and the local encoding Q_{acc}' innervates the ocular muscles for accommodation, so that the optical system focuses the messenger light of the fixation point on the fovea. Eventually, the observer sees a sharp image Q'' behind the mirror.

Fig. 5



Separating visual projection from its original by regulating the afferent messenger-light pathway with a mirror. **a**, A mirror is fixed on one side of a sealed metal box, whereas a ball Q , covered by a semi-enclosed barrier hangs above the observer. The lamp, when turned on, fires detective photons whose messenger-light feedforward signaling pathway (FSPm) from the unseen object Q to the retina of the observer is divided into two segments, S_1 and S_2 . The messenger light in S_2 induces the retina to fire action potentials (APs) (Q'). Q' enters the AP feedforward signaling pathway (FSPa), from which the top-down encodings Q_{mov}' (not shown, see Fig. 4a for detail) and the local encoding Q_{acc}' arise to regulate the optical system (Kandel et al. 2012). As a result, the messenger light of the fixation point (the dark brown centerline of S_1 and S_2) is focused on the fovea, and the messenger light of Q is focused collaterally on the retina. Eventually, a sharp reflection Q'' of Q appears behind the mirror. **b**, After the metal box and mirror are withdrawn, the observer watches the ball again (here, the original Q is not shown; see Fig. 4c for detail).

Let us now withdraw the mirror and box and move the ball to the place where the preceding Q'' used to be, while leaving the rest of the scenario unchanged (Fig. 5b). When the observer fixates on the current fixation point (i.e., the previous virtual fixation point), the terminal pathway of the current FSPm, S_2' is in line with the previous S_2 ; accordingly, the top-down encoding Q_{mov}' and local encoding Q_{acc}' for the current FSPm are the same as those for S_2 . The same is true for the intermediate- and high-level visuospatial processing. Consequently, in either case, the NCC projects the same image to the same place; i.e., both observation modes are equivalent for visual perception. To distinguish the two types of OBP, the OBP of brain-generated imagery spatially superimposed onto its original is termed classic projection, whereas other OBPs of brain-generated imagery, such as starlight deflection, mirror image, afterimage, imagination, and dreams, are termed nontrivial projections. More importantly, in contrast to the classic projection, the mirror image clearly shows that

$$\forall(t\xi)(s(t\xi) \cap s''(t\xi) = \emptyset),$$

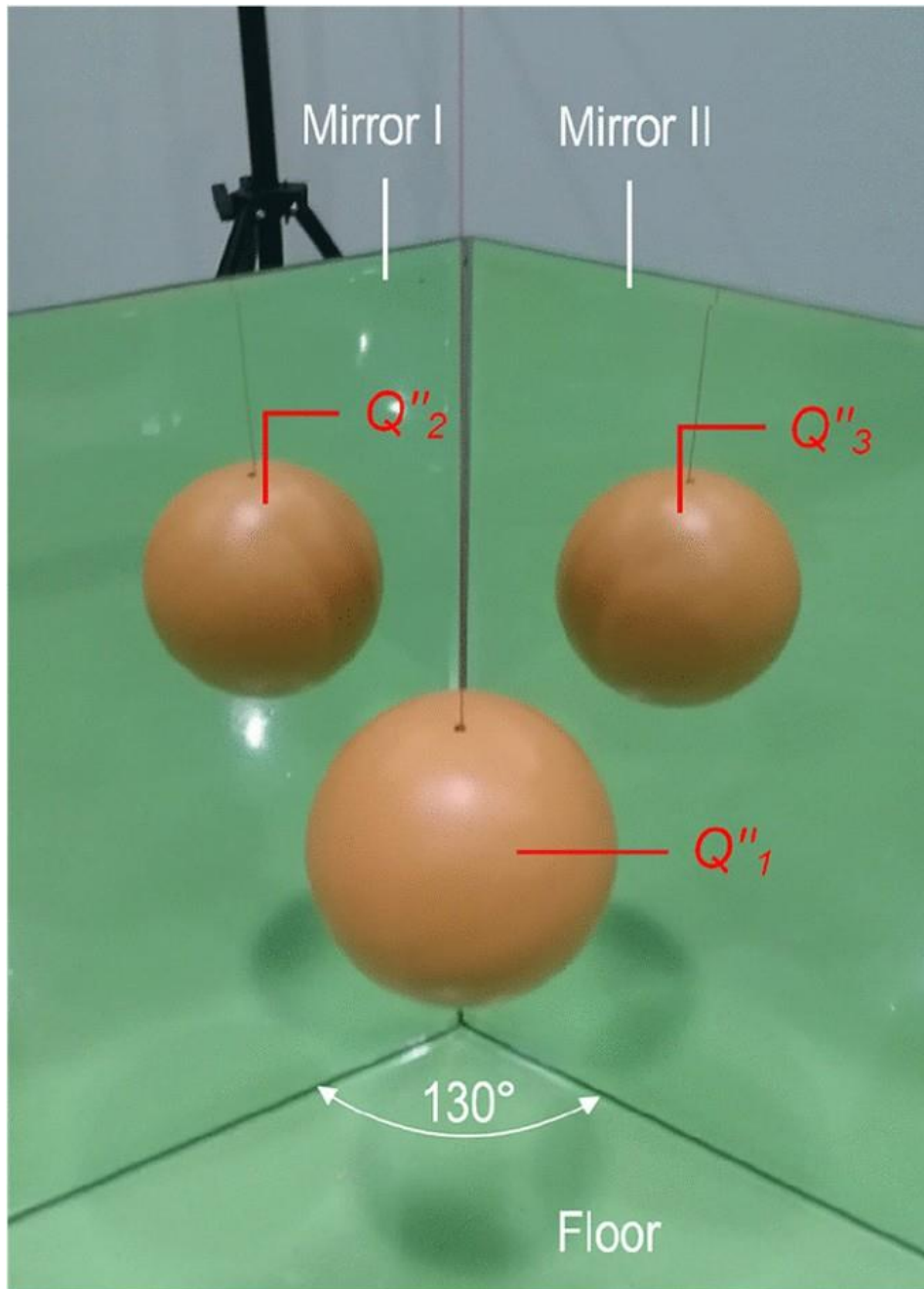
where s and s'' are the spaces owned separately by Q and Q'' , and t_ξ is the time variable that traverses throughout the observation period. Based on the non-identity law, we

obtain $Q'' \cap Q = \emptyset$, revealing that both still objects, the visual projection Q'' and the original Q , are two different things. It is noteworthy that Q'' and Q are deemed to be symmetric with respect to the mirror; however, this cannot be experimentally determined. For this reason, a thought experiment is adopted instead of a practical experiment.

Moreover, the mirror image provides evidence that the OBP pathway, in response to the messenger light of the original, has constancy, and that the OBP can be manipulated by regulating or reconstructing the FSPm. The FSPm-regulating technique, whereby FSPm is regulated in real time has long been used to create optical tools, such as the microscopes, telescopes, periscopes, and spectacles, whereas the FSPm-reconstructing technique, whereby FSPm can be reconstructed anytime and anywhere, is used to make image and video productions, such as 3-dimensional paintings and movies, and virtual reality. Additionally, the FSPm-regulating technique can be used to treat chronic neurological disorders, such as phantom limb pain and hemiparesis caused by stroke (Ramachandran & Altschuler, [2009](#)). It can even be used to unveil a nontrivial mental phenomenon, the out-of-body experience (Ehrsson, [2007](#); Lenggenhager et al., [2007](#)). In summary, these seemingly isolated manifestations are just different forms of nontrivial projections whose FSPm are regulated, reconstructed, or simulated.

To understand the classic and nontrivial projections further, we constructed a scene where one ball and two mirror images appeared simultaneously (Fig. [6](#)), prompting the question: which one is true? Based on the OBP principle, the ball that could be touched was considered a classic projection, whereas the two other mirror images were nontrivial projections; therefore, all of them were true images but none of them was the real object^{[Footnote4](#)}. Based on the non-identity law, the three images located at different positions were three things. There is, however, little doubt that all of them, which were very much alike in appearance, corresponded to the same unknown original. This demonstrates that an original can simultaneously present multiple images to an observer.

Fig. 6



Multi-image perception of an original. The test equipment comprised two mirrors (56×35 cm, L $\times$ W) and an orange metal ball with a diameter of 15 cm. The mirrors are close to each other and have an intersection angle of approximately 130 degrees. The ball is suspended approximately 25 cm above the floor. $Q1''$ is a “ball” that can be seen and touched simultaneously, whereas $Q2''$ and $Q3''$, which cannot be touched, are two nontrivial images whose messenger-light feedforward signaling pathways are regulated by mirror I and mirror II, respectively. The three images possess the same shape and color but differ in their positions and directions. Note that $Q1''$, which could be touched by an observer, is generally regarded as a real object; however, it is merely a classic projection of brain-generated imagery in response to the messenger light of an unknowable original.

Discussion

Is visual projection a physical behavior?

Scholars have long been aware of the OBP of an afterimage (the afterimage can be projected anywhere and the size increases with the projection distance), and have proposed Emmert's law (Boring, [1940](#)), which provides a nontrivial cue for revealing visual awareness. Perhaps, because subsequent research has focused on the feedforward size constancy instead of visual projection (Epstein et al., [1961](#)), and there is an unexplainable deviation occurring between the hypothesis and fact (Young, [1948](#); Furedy & Stanley, [1970](#)), the visual projection has been regarded as an isolated visual phenomenon and has not been developed into a universal visual principle, until now. The root cause, however, may lie within consciousness itself.

Consciousness, generally considered the subjective experience arising from the brain, has several typical components, such as sensation, emotion, reasoning, imagination, and self-awareness, which possess some common features, such as subjectivity and intentionality (Searle, [1984](#), [2004](#)). Then, what about visual awareness as a representative sensation?

First, visual awareness is based on an FSP-OBP pathway whose FSP has a physical delay of approximately 100 ms after flash onset. Similarly, related research on motion control has demonstrated that a decision can be encoded into the prefrontal and parietal cortex up to 10 seconds before it enters awareness (Soon et al., [2008](#)). Furthermore, a lesion to the FSPa may pose a loss of conscious outcome, such as blindsight and motion blindness (Lamme, [2001](#); Goodale & Milner, [2013](#); Zihl et al., [1983](#)), resulting in a remarkable deterioration in behavioral flexibility. However, even though the FSP functions normally, all brain activities do not cause conscious awareness (Custers & Aarts, [2010](#)). These provide converging evidence that the postponed conscious outcomes of sight and motion are determined by sophisticated brain activity, supporting the argument that conscious outcome is neither predictive nor online, but rather postdictive (Eagleman & Sejnowski, [2000](#)). It also supports the hypothesis that conscious outcome is a collateral product or epiphenomenon of brain activity (Huxley, [1874](#)), similar to a host-controlled screen that selectively displays something. Although this view is not widely accepted, this principle is used to conduct neurological investigations into and clinical treatments for patients with cognitive disorders (e.g., Alzheimer's and Parkinson's disease) and to anesthetize a patient before an operation to deprive the patient of pain and even awareness.

Second, the media of the FSP ($pd \rightarrow Q \rightarrow prRetina \rightarrow Q'NCC$) are known to be photons and APs in physical terms. Analogously, we assume that the medium of the OBP ($NCC \Rightarrow Q''$) is x ; hence, the updated OBP expression is $NCC \Rightarrow xQ''$. This prompts the following question: what is x ? The materials that support life are the same as those present in the universe. Since the operation of an object can be influenced by its interaction with other objects (Ridley [1995](#)), we can always use certain materials to intervene with the OBP if the visual projection is a physical behavior, as we have done in other scientific experiments. However, we have never encountered a situation where a wall (or a sealed metal box, or anything else) on which a mirror is fixed

blocks the OBP pathway and obstructs the visual perception of the mirror image. Moreover, the color of the object, whether it is inside or outside the mirror, cannot be deduced from the three fundamental concepts in physics. Therefore, a mirror image is not a physical thing and OBP is not a physical behavior, thereby falling into the so-called phenomenological or transcendental category (Kant, [1783](#)). It is worth emphasizing that the method can be called a mirror test, which is likewise suitable for examining theories of consciousness.

Third, it should be highlighted that the brain-generated imagery is projected back onto an original in response to the afferent messengers of the original, signifying that the NCC can perform voluntary OBP, whose behavior can be termed projection intentionality. This evidence indicates that humans have an instinct for producing subjective OBP to keep up with its original in space, like a searchlight trying to track and shine on a moving object in the dark, or a bat positioning and catching a moth in the dark. The same is true for other sensations, such as touch and hearing; for example, when we touch something with a tool, we can feel the touch at the tip of the tool, as an extension of the body (Miller et al., [2018](#); Heed, [2019](#)), signifying that touch also possesses the capacity for OBP, extending beyond the boundary of the body. Accumulating evidence has demonstrated that human sensations, such as vision, touch, and hearing, possess the attribute of spatial expansion outward from the body. Thus, it can be inferred that intentionality is one aspect of the subjectivity of consciousness and that the recurrent OBP may be the psychophysical origin of the intentionality of consciousness.

Taken together, all experiences, the responses to the messengers of the originals or to their memory recall, should fall into the subjective manifestation category. For example, “space,” “time,” and “acting force,” generally regarded as so-called objective things, are merely the measurable components of conscious outcomes that can be reciprocally verified by different perceptions (particularly, visual and tactile perceptions), suggesting that what one perceives is simply a fantasy that coincides with reality; i.e., none of the experiences are original but are merely subjective perceptions that metaphysically reflect particular originals via the afferent messengers of the originals^{[Footnote5](#)}. Hence, it is not difficult to understand why the symmetry of the mirror image has never been experimentally determined and why the visual projection presented in afterimages has not been developed into a universal visual principle.

Isomorphic relations between original, experience, and expression

Thus far, a panorama of the universe has unfolded. The universe for an observer (first-person) is composed of three components (Fig. [7](#)): two worlds and one observer (i.e., perceiver).

1. (1)

The originals’ world in which some originals act as the messengers of other originals and can be perceived by the observer.

2. (2)

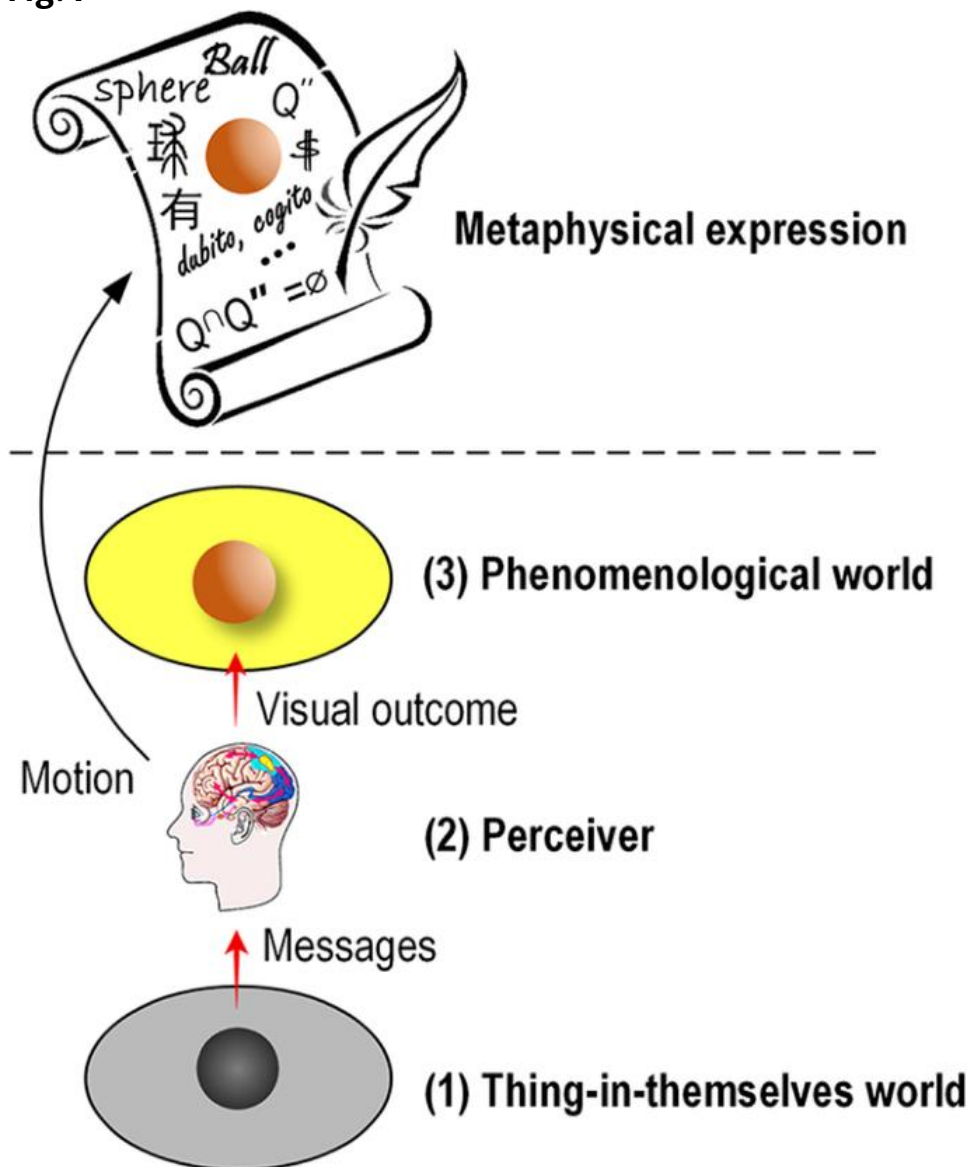
The observer in which two transformations occur. The observer’s body is a part of the originals’ world, and hence it can interact with the originals. Unfortunately, the

unconscious neural activity responding to an afferent sensory stimulus or memory recall is not perceived by any perceiver; fortunately, we possess the instinct of not only subjectively imaging but also projecting the image back onto the original to form subjective perception. From the perspective of information flow, the perceiver is a transparent^{Footnote6} component between the originals' world and the phenomenological world.

3. (3)

The phenomenological world (i.e., conscious outcome, also known as subjective perception, experience, phenomenon, imagery, appearance, manifestation, or representation) that follows or reflects the dark originals. It is well known that the subjective perception is private, which is the fact claimed by Wittgenstein.

Fig. 7



Three-component world and its metaphysical expression.

Furthermore, the metaphysical representation (expression) manipulated by both human memory and human motor system describes the facts presented in the phenomenological world, is a unique achievement of humans — in this metaphysical field, Wittgenstein spent his whole life exploring the inherent logic of language consisting of symbols, mathematical expressions, words and pictures. The metaphysical expression is important, and therefore, it is also included here to render the discussion complete.

The three-component world view corroborates the following well-known philosophical ideas: “although the thing is completely unknown to us as to what they may be in themselves, we know through the representations which their influence on our sensibility, and to which we give the name of a body” (Kant, [1783](#)); “manifestation stems from thing-in-itself. ... thing-in-itself and its manifestation are the same, though they are named differently, which are called mysteries, and the mystery underlying the mysteries is the gateway to all understanding” (Lao-Tzu, 480 BC). Lao-Tzu further argued that constant manifestations can be used to investigate the bounds of a thing, which coincides with the basis of the non-identity law. Interestingly, this argument is supported by the discovery that neurons performing visuospatial processing at different levels are most sensitive to the borders and contours of the observed objects in a scene (Kandel et al., [2012](#)).

However, not all messengers can induce sensory receptors to shape an experience. For example, invisible originals, such as air and most types of radiation, provide evidence that in some cases, human sensory experience is absent from the operation of varied originals^{[Footnote7](#)}. Additionally, the nontrivial projection mentioned above demonstrates that the experience may deviate markedly from its original. Therefore, human experience is a limited and even nontrivial reflection of the originals in some cases. It must be emphasized that a sensory experience, i.e., the response to the messengers of a particular original, is trustworthy, yet also questionable.

Since an experience, which arises from the NCC and fades as the messengers disappear, is a transient, private, and subjective event, how do we record and express our feelings and communicate with each other? We have to use certain languages (gestures, sounds, characters, symbols, or drawings) that can be perceived by others using sight, hearing, and touch, to describe, name, and record the phenomenological events, and statistically develop a “constant conjunction” between different events (Hume, [1748](#)). This process is labeled phenomenological expression, which is a man-made, abstract, and limited mapping (isomorphism) of the experience. As such, it is replete with symbolization and suggestiveness. Interestingly, in turn, the symbolic and kinematic suggestiveness of the expression leaves something to our imagination, resulting in several rigorous metaphysical systems, such as mathematics, logic, and graphic art, which have laid the groundwork for scientific research.

We can go even further; using the above metaphysical systems, we can consolidate as many manifestations within a unified metaphysical framework as possible. This process may be termed metaphysics-consolidated expressions, such as conservation laws, evolution theory, Maxwell's

equations, special relativity, and the standard model. In parallel, through the phenomenological observations by the aid of advancements in experimental apparatus combined with proper rational analysis, scientists have made many significant empirical findings such as elementary elements, elementary particles, the constant speed of light, and the DNA double-helix structure.

In comparison with the phenomenological and metaphysics-consolidated expressions, it is more revolutionary to seek the cause of the manifestations. First, regarding certain manifestations as evidence, we tentatively put forward a hypothesis as the common cause for the manifestations and model it with several acknowledged constant laws, such as interaction laws, conservation laws, and evolution rules. Next, if the deduction or laboratory development on the hypothesis does not fit into the current and hypothetical-deductive-forecasted evidence, we revise the hypothesis until it fits the evidence well. Subsequently, the best-matching hypothesis is regarded to be true. This process is labeled hypothetical expression. Interestingly, it is the hypothesis-to-manifestation research approach that has helped us break through cognitive limitations and that fuels the progress of science, resulting in a series of theories, such as heliocentrism, gravity, atomic model, energy quantum, and general relativity.

However, both phenomenological and hypothetical expressions remain tentative. A manifestation is only a limited or even a nontrivial reflection of the real world; therefore, the phenomenological expressions extracted from the manifestations are fallible, which further affects the legitimacy of metaphysics-consolidated expressions. The hypothesis-to-manifestation expressive paradigm may put our cognition or metaphysics at risk of straying from the natural original-to-manifestation route. It is, therefore, inevitable and unsurprising that multiple hypotheses related to the same issue may exist in parallel for a long time. Unfortunately, we have no other choice^{[Footnote8](#)}. These provide insight into the core thought of Taoism: “we can name and describe manifestations, think and talk about Tao underlying manifestations, but the expressions are not manifestations and Tao themselves and, consequently, are fallible; ... nevertheless, we can still explore the mystery of Tao and the bounds of manifestations by their constancy” (Lao-Tzu, 480 BC).

Since manifestations and expressions are fallible, how do we ensure that our intellectual adventure is safe? In addition to rigorous algorithmic rules, the expressed hypotheses or truths should satisfy the criteria of compatibility, completeness, and simplicity. As far as the mind-body problem is concerned, Kant’s transcendental philosophy and Lao-Tzu’s Wu-You thought, which acknowledge the existence of unknowable things in themselves, are not only compatible with each other but are also applicable to classic and nontrivial imaging without recruiting free will. They are thoroughly materialistic philosophies^{[Footnote9](#)} (in fact, most of us usually see the things before our eyes as real things, and hence, are idealists). Contrastingly, parallel traditional philosophies, such as dualism, mechanical materialism, and idealism, which treat free will as the cause of human activities to varying degrees, are neither compatible with each other nor able to pass the mirror test (that is to say, none of these traditional philosophies are able to explain the phenomena shown in Fig. [5a](#) and Fig. [6](#)). Fortunately, subjective perception generally conforms to its original. However, this is not always the case. Therefore, we should be open to the

expressions with respect to the manifestations and their underlying causes, including the expressions in this paper.

Conclusion

The current study primarily sought to address what consciousness consists of from the angle of visual awareness. It consists of a comprehensive investigation of the global visual dynamics between the thing-in-itself, brain, and visual experience, via an interdisciplinary approach involving philosophy, physics, logic, neuroscience, and psychophysics. The results revealed that visual awareness involves a postponed, recurrent OBP, suggesting that the visual system has an instinct of not only subjectively imaging, but also projecting the image back onto its original, according to the cue of the afferent messenger-light pathway of the original. This finding, coupled with the NCC, adds a key puzzle piece to the visual system, forming a complete visual causal chain. Contrastingly, a lack of OBP may result in a remarkable deterioration in behavioral flexibility, such as blindsight and motion blindness. A possible explanation for this may be that the OBP is an optimum evolutionary strategy and is essential for the survival of fast-moving creatures.

In light of this finding, this study clearly explained nontrivial projections such as the mirror image, suggested the mirror test as a criterion for examining theories of consciousness, elucidated the psychophysical root of both subjectivity and intentionality of consciousness, and highlighted a growing principled understanding of the isomorphic relations between the original, experience, and expression. These results suggest that under adopted consciousness, the interdisciplinary integration of cognitive science can be further promoted and that the debate regarding the bounds of artificial intelligence, i.e., whether a machine can have consciousness, may be settled. Although OBP plays a crucial role in visual awareness, the recurrent OBP pathway is only roughly provided in this study, while the only related hypothesis, Emmert's law, is at variance with the fact. Therefore, the projection geometry of the OBP remains to be experimentally determined. The solution to the projection geometry will further prove that consciousness, including self-awareness, can be understood in physical terms. Further research in this field will require the interpretation of another aspect of the “hard problem” — how consciousness arises from the brain activity of life developed from a fertilized egg, and empirical studies that address issues regarding more accurate dynamics of sensing, emotion, and thinking.

Data Availability

All data is available in the main text or in the supplementary material, and the materials required in the current study, including the schematic and PCB of the specially designed circuit for the experiment, are available from the author on reasonable request.

Notes

1. It should be heeded that there are only two real things in the model of boundary of A for simplicity. In fact, there are various things around A in real world. When we conduct

empirical research on A, around which some things are accidentally or intentionally ignored, e.g., the extremely faint interaction between A and distant (known or unknown) celestial bodies, indicating that the empirical study is always incomplete. Therefore, the ICP is a universal principle presented in the scientific studies. Furthermore, if only research is conducted on universe A itself, then the research on universe A is also incomplete, signifying the limited ability of man exploring the mysteries of the universe.

2. The absolute non-identity law allows us to distinguish one thing from another, depending on their constancy and observability (appearances and bounds) in the space dimension, providing a bridge from the appearance of an object to the concept of the object. From the perspective of the knowledge system, the law of non-identity is the premise of all human cognition, including the law of excluded middle and the law of contradiction. If nothing is distinguishable in human perception, there is only one thing remaining, called “chaos.”
3. Any substance takes time to move from one position to another. Its speed cannot exceed the speed of light ($c_0=299792458$ m/s). Here, the travel of light also requires time, but due to the short transmission distance, the transmission time of light is negligible. In comparison, the duration of light-to-electricity transformation and the period of action potentials traveling from one site to another are significant.
4. All the images are regarded as “true,” because perception is the starting point of human cognition; i.e., man is the yardstick of everything. In other words, what one perceives, even a mirage, is a “true” perception. Notably, “true” image does not mean that its original actually exists, e.g., computer-generated imagery.
5. Interestingly, this article began with the law of non-identity achieved by investigating the condensed matter represented in the time and space, but ultimately resulted in the query to the objectivity of time and space. Time, space, and matter are not only scientific issues but also classic philosophical issues. These issues concern the nature of the universe and its motion. “The flying arrow is at rest,” a well-known paradox proposed by Zeno (Huggett, [2019](#)), is yet to be resolved systematically and quantitatively.
6. ‘transparent’, a terminology in the computer field, is used here to indicate that a brain is like a super Black Box — it is impossible for us (third-person) to observe the private subjective perception of any participant or to infer the one using scientific laws and theories. Brain is the bridge from originals’ word to the phenomenological world and undoubtedly the most mysterious thing in the universe.
7. Each species has its own limited but sufficient sentience for its survival, which is the inevitable result of evolution. The same is true of the selective perception of humankind. Too much sentience would otherwise only bring about confusion and waste. Even if one species historically had other forms of sentience, lack of use usually resulted in the degradation of that sentience, for example the visual degradation of moles.
8. In his Hugo Award-winning science fiction novel "The Three-Body Problem" (2015), Cixin Liu vividly portrays the dilemma faced by the Trisolarans in explaining complex

and ever-changing celestial phenomena, which is a true reflection of humanity's own struggles with similar dilemma.

9. Things in themselves refer to the dark unknowable things in the objective originals' world that underlies the subjective phenomenological world, including the dark things underlying the perceptions of matter, space, and time, as well as the interaction between different matters. Literally, therefore, the word "materialistic" is an incomplete makeshift expression and should be completed in the future.

References

- Albright, T. D., Jessell, T. M., Kandel, E. R., & Posner, M. I. (2000). Neural science: A century of progress and the mysteries that remain. *Cell*, 100, S1–S55. [https://doi.org/10.1016/S0896-6273\(00\)80912-5](https://doi.org/10.1016/S0896-6273(00)80912-5)

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Batista, A. P., Buneo, C. A., Snyder, L. H., & Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science*, 285(5425), 257–260. <https://doi.org/10.1126/science.285.5425.257>

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Bayes, T. ([1763] 2003). An essay towards solving a problem in the doctrine of chances. *Resonance*. <https://doi.org/10.1007/BF02883540>
- Bear, M. F., Connors, B. W., & Paradiso, M. A. (2015). *Neuroscience: Exploring the Brain* (4th ed.). Lippincott Williams & Wilkins.
- Blackmore, S. J. (2005). *Consciousness: A Very Short Introduction*. Oxford University Press.

[Book](#) [Google Scholar](#)

- Blackmore, S. J. (2006). *Conversations on Consciousness: What the Best Minds Think About the Brain, Free Will, and What It Means to Be Human*. Oxford University Press.

[Google Scholar](#)

- Boring, E. G. (1940). Size Constancy and Emmert's Law. *American Journal of Psychology*, 53, 293–295. <https://doi.org/10.2307/1417427>

[Article](#) [Google Scholar](#)

- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.

[Google Scholar](#)

- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, D. J. (2013). How can we construct a science of consciousness? *Annals of the New York Academy of Sciences*, 1303, 25–35. <https://doi.org/10.1111/nyas.12166>

[Article ADS PubMed Google Scholar](#)

- Chen, J., Wu, E.-D., Chen, X., Zhu, L.-H., Li, X., Thorn, F., Ostrovsky, Y., & Qu, J. (2016). Rapid integration of tactile and visual information by a newly sighted child. *Current Biology*, 26, 1069–1074. <https://doi.org/10.1016/j.cub.2016.02.065>

[Article CAS PubMed Google Scholar](#)

- Clifford, W. K. (1886). *The common sense of the exact science* (2nd ed). Kegan Paul, Trench, & Co
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275. <https://doi.org/10.1038/nn0203-119>

[Article CAS Google Scholar](#)

- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6, 119–126. <https://doi.org/10.1038/nn0203-119>

[Article CAS PubMed Google Scholar](#)

- Custers, R., & Aarts, H. (2010). The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science*, 329, 47–50. <https://doi.org/10.1126/science.1188595>

[Article ADS CAS PubMed Google Scholar](#)

- Descartes, R. ([1637] 1985). *Discourse on the Method of Rightly Conducting the Reason and Seeking Truth in the Sciences*. In J. Cottingham, R. Stoothoff & D. Murdoch (trans.), *The Philosophical Writings of Descartes (1)*. Cambridge University Press
- Dennett, D. C. (1991). *Consciousness Explained*. Little Brown & Co..

[Google Scholar](#)

- Dobbins, A. C., Jeo, R. M., Fiser, J., & Allman, J. M. (1998). Distance modulation of neural activity in the visual cortex. *Science*, 281, 552–555. <https://doi.org/10.1126/science.281.5376.552>

[Article ADS CAS PubMed Google Scholar](#)

- Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, 287, 2036–2038. <https://doi.org/10.1126/science.287.5460.2036>

[Article ADS CAS PubMed Google Scholar](#)

- Eccles, J. C. (1951). Hypotheses relating to the brain–mind problem. *Nature*, 168, 53–57. <https://doi.org/10.1038/168053a0>

[Article ADS CAS PubMed Google Scholar](#)

- Edelman, G. M. (1989). *The Remembered Present: A Biological Theory of Consciousness*. Basic Books.

[Google Scholar](#)

- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, 317, 1048. <https://doi.org/10.1126/science.1142175>

[Article ADS CAS PubMed Google Scholar](#)

- Epstein, W., Park, J., & Casey, A. (1961). The current status of the size-distance hypotheses. *Psychological Bulletin*, 58, 491–514. <https://doi.org/10.1037/h0042260>

[Article CAS PubMed Google Scholar](#)

- Furedy, J. J., & Stanley, G. (1970). The apparent size of “projected” afterimages under conditions where size-constancy holds. *Perception & Psychophysics*, 7, 165–168. <https://doi.org/10.3758/BF03208650>

[Article Google Scholar](#)

- Gentile, G., Petkova, V. I., & Ehrsson, H. H. (2011). Integration of visual and tactile signals from the hand in the human brain: An fMRI study. *The Journal of Neurophysiology*, 105, 910–922. <https://doi.org/10.1152/jn.00840.2010>

[Article PubMed Google Scholar](#)

- Gentzen, G. ([1938] 1969). The present state of research into the foundations of mathematics. in M. E. Szabo (trans. and ed.), *The Collected Works of Gerhard Gentzen*. North-Holland Publishing Co.
- Gentzen, G. (1969). The consistency of elementary number theory. in M. E. Szabo (trans. and ed.), *The Collected Works of Gerhard Gentzen*, North-Holland Publishing Co.
- Goodale, M. A., & Milner, D. (2013). *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. Oxford University Press.

[Book Google Scholar](#)

- Hammersley, J. M., & Handscomb, D. C. (1965). *Monte Carlo Methods*. Fletcher & Son Ltd..

[Google Scholar](#)

- Healy, K., McNally, L., Ruxton, G. D., Cooper, N., & Jackson, A. L. (2013). Metabolic rate and body size are linked with perception of temporal information. *Animal Behaviour*, 86, 685–696. <https://doi.org/10.1016/j.anbehav.2013.06.018>

[Article PubMed PubMed Central Google Scholar](#)

- Heed, T. (2019). Tool use: two mechanisms but one experience. *Current Biology*, 29, R1301–R1303. <https://doi.org/10.1016/j.cub.2019.10.062>

[Article CAS PubMed Google Scholar](#)

- Huggett, N. (2019). Zeno's Paradoxes. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2019/entries/paradox-zeno/>
- Hume, D. ([1748] 2007). *An Enquiry concerning Human Understanding*, P. Millican (ed.). Oxford University Press.
- Huxley, T. ([1874] 2012). On the hypothesis that animals are automata, and its history. In *Collected Essays: Methods and Results 1* (pp. 199-250). Cambridge University Press. <https://doi.org/10.1017/CBO9781139149204>
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2012). *Principles of Neural Science* (5th ed.). McGraw-Hill Medical.
- Kant, I. ([1783] 2004). *Prolegomena to Any Future Metaphysics, That Will Be Able to Come Forward as Science with Selections from the Critique of Pure Reason*, G. Hatfield (ed.). Cambridge University Press.
- Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Company Publishers.

[Google Scholar](#)

- Koch, C., & Hepp, K. (2006). Quantum mechanics in the brain. *Nature*, 440, 611–612. <https://doi.org/10.1038/440611a>

[Article ADS CAS PubMed Google Scholar](#)

- Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, **17**, 307–321. <https://doi.org/10.1038/nrn.2016.22>

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, **23**, 571–579. [https://doi.org/10.1016/S0166-2236\(00\)01657-X](https://doi.org/10.1016/S0166-2236(00)01657-X)

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Lamme, V. A. F. (2001). Blindsight: The role of feedforward and feedback corticocortical connections. *Acta Psychologica*, **107**, 209–228. [https://doi.org/10.1016/S0001-6918\(01\)00020-8](https://doi.org/10.1016/S0001-6918(01)00020-8)

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Lao-Tzu. ([circa 480 BC] 1996). Tao Te Ching. In G. Ming (ed.), *Variorum of Silk Book of Laozi* (Chap. 1 & 40). Zhonghua Book Co.
- Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, **317**, 1096–1099. <https://doi.org/10.1126/science.1143439>

[Article](#) [ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, **64**, 354–361. <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>

[Article](#) [Google Scholar](#)

- Mashour, G. A. (2018). The controversial correlates of consciousness. *Science*, **360**, 493–494. <https://doi.org/10.1126/science.aat5616>

[Article](#) [ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Michael, A. C. (2015). The problem with Phi: A critique of integrated information theory. *PLOS Computational Biology*, **11**(9), e1004286. <https://doi.org/10.1371/journal.pcbi.1004286>

[Article](#) [CAS](#) [Google Scholar](#)

- Miller, L. E., Montroni, L., Koun, E., Salemme, R., Hayward, V., & Farne, A. (2018). Sensing with tools extends somatosensory processing beyond the body. *Nature*, **561**, 239–242. <https://doi.org/10.1038/s41586-018-0460-0>

[Article](#) [ADS](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neuroscience*, **6**, 414–417. [https://doi.org/10.1016/0166-2236\(83\)90190-X](https://doi.org/10.1016/0166-2236(83)90190-X)

[Article Google Scholar](#)

- Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, **3**, 82–791. <https://doi.org/10.1038/s41562-019-0626-2>

[Article Google Scholar](#)

- Ramachandran, V. S., & Altschuler, E. L. (2009). The use of visual feedback, in particular mirror visual feedback, in restoring brain function. *Brain*, **132**, 1693–1710. <https://doi.org/10.1093/brain/awp135>

[Article CAS PubMed Google Scholar](#)

- Ridley, B. K. (1995). *Time, Space and Things* (3th ed.). Cambridge University Press.
- Schwartz, E. L., Desimone, R., Albright, T. D., & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences*, **80**(18), 5776–5778. <https://doi.org/10.1073/pnas.80.18.5776>

[Article ADS CAS Google Scholar](#)

- Searle, J. R. (1984). *Minds, Brains and Science*. Harvard University Press.
- Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences*, **16**, 585–642. <https://doi.org/10.1017/S0140525X00080304>

[Article Google Scholar](#)

- Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press.

[Book Google Scholar](#)

- Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, **11**, 543–545. <https://doi.org/10.1038/nn.2112>

[Article CAS PubMed Google Scholar](#)

- Tapia, E., & Beck, D. M. (2014). Probing feedforward and feedback contributions to awareness with visual masking and transcranial magnetic stimulation, *Frontiers in Psychology*, **5**, Article 1173. <https://doi.org/10.3389/fpsyg.2014.01173>

- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6, 171–178. [https://doi.org/10.1016/S0959-4388\(96\)80070-5](https://doi.org/10.1016/S0959-4388(96)80070-5)

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Van Gulick, R. (2018). Consciousness. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018.). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2018/entries/consciousness/>
- Walsh, V., & Cowey, A. (1998). Magnetic stimulation studies of visual cognition. *Trends in Cognitive Sciences*, 2, 103–110. [https://doi.org/10.1016/S1364-6613\(98\)01134-6](https://doi.org/10.1016/S1364-6613(98)01134-6)

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Wittgenstein, L. ([1921] 2001). *Tractatus Logico-Philosophicus*. Routledge.
- Young, F. A. (1948). The projection of after-images and Emmert’s law. *The Journal of General Psychology*, 39, 161–166. <https://doi.org/10.1080/00221309.1948.9918172>

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

- Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain*, 106, 313–340. <https://doi.org/10.1093/brain/106.2.313>

Research on How Human Intelligence, Consciousness, and Cognitive Computing Affect the Development of Artificial Intelligence

[Yanyan Dong](#), [Jie Hou](#), [Ning Zhang](#), [Maocong Zhang](#)

First published: 28 October 2020

<https://doi.org/10.1155/2020/1680845>

Abstract

Artificial intelligence (AI) is essentially the simulation of human intelligence. Today's AI can only simulate, replace, extend, or expand part of human intelligence. In the future, the research and development of cutting-edge technologies such as brain-computer interface (BCI) together with the development of the human brain will eventually usher in a strong AI era, when AI can simulate and replace human's imagination, emotion, intuition, potential, tacit knowledge, and other kinds of personalized intelligence. Breakthroughs in algorithms represented by cognitive computing promote the continuous penetration of AI into fields such as education, commerce, and medical treatment to build up AI service space. As to human concern, namely, who controls whom between humankind and intelligent machines, the answer is that AI can only become a service provider for human beings, demonstrating the value rationality of following ethics.

1. Introduction

The term "artificial intelligence" was first used by John McCarthy at Dartmouth Conference in 1956. Since then, artificial intelligence (AI) has gone through three booms during decades of scientific and technological development. The first boom was from 1956 to 1976. Since the 1950s, humans had successively invented the first perception neural network software and chat software and proved some mathematical theorems, exclaiming that the "AI era is coming" and that "robots will surpass human beings in 10 years." During the second boom (1976–2006), Hopfield neural network [1] and BT training algorithm proposed in the 1980s made AI popular again, which led to the emergence of speech recognition, speech translation plan, and Japan's fifth-generation computer idea. However, these ideas fell through and the second boom broke up again. After data accumulated to a certain amount, some results would stop rising to some extent. During the third boom (2006 till now), AI broke out again as Hinton put forward deep learning technology in 2006 and ImageNet Competition made breakthroughs in image

recognition in 2012. In 2016, AlphaGo defeated Lee Se-dol, once the world champion of Go, and that was regarded as the peak of AI development.

Now, humans have made great progress in various fields such as cognitive psychology, neuroscience, quantum physics, and brain science, and theories related to artificial intelligence have kept emerging. Without the integrated development of computer science with brain science, neuropsychology, linguistics, and other disciplines, the research and development of AI would not have made such great achievements. AI research has also presented some research highlights such as machine learning, neural network (NN), expert system, genetic algorithm (GA), fuzzy inference system (FIS), support vector machine (SVM), and particle swarm optimization (PSO) [2], as shown in Figure 1. AI has been widely used in each aspect of human life, even surpassing human intelligence in some areas. AI can, to some extent, replace humans to complete the tasks of recognition, decision-making, and control. In terms of recognition, AI can distinguish, classify, and retrieve information. In the aspect of decision-making, AI can carry out numerical object evaluation and matching. Regarding control, AI can complete performance generation, design and action optimization, and operation automation.

Figure 1

[Open in figure viewerPowerPoint](#)

Knowledge graph of AI research hotspots.

However, scientific, technological, and social problems arisen from the development of AI are drawing more attention from the public. While promoting social progress, the widespread application of AI also has some prominent negative effects. For example, the application of robot has led to unemployment [3]; the application of AI has widened the gap of wealth; AI algorithms have caused bias; big data has resulted in privacy leakage and degeneration of human's spiritual life. All these are social problems brought about by the application of AI technologies. In his book *AI Future*, Kai-fu Lee analyzed the employment risks of manual and mental labor caused by AI. For manual workers, structured jobs with weak social interactions and low skills are facing higher risks, such as truck drivers, fast food cooks, and sewing workers. For mental workers, those engaged in weak social and low creative jobs are in greater danger of unemployment, such as radiologists and telemarketers. Therefore, in the AI era, it is inevitable for some occupations or jobs to be replaced. If your job does not require much talent and skills can be acquired through training, if what you need to do is a lot of repetitive work without too much thinking, and if you work in a narrow social network and have little communication with others, then you are highly likely to be replaced by AI.

The deconstruction and reconstruction of the occupational structure is just one aspect of what AI has brought about. Some people even put forward the "machine threat theory." Understanding AI is the prerequisite of applying it. AI is in essence the simulation of human intelligence and its development depends on breakthroughs of algorithms. Considering this relationship as well as human's confusion and concerns in understanding AI, this paper first

analyzes the nature of AI, AI, and consciousness, as well as the features and advantages of cognitive computing, and further forecasts the future development of AI.

2. The Nature and Characteristics of AI

The difference between AI and ordinary computers lies in the new forms and technical means of AI, but they share the same nature. AI is still a tool of humans [4]. To accurately grasp AI, we can make an analysis from two aspects: hardware and software. From the hardware point of view, intelligent machines like computers and robots are physical and chemical entities separated from human brains. Although they are not physiological structures of human reflection and control systems, they are developed from the hardware of the system. The human brain is also hardware, but it is a physiological entity instead of a physical or chemical one. The human brain provides the material base of the reflection and control function. In this sense, the brain entity provides the structure and thinking offers the function. Intelligent machines have physical and chemical structures, in which electrons make a physical movement. This structure is a prerequisite for AI. The physiological structure of the human brain is different from the structure of intelligent machines, so they have no physiological or fixed hardware connections. Therefore, although intelligent machines are developed from the hardware of human's reflection and control system, their development is relatively independent. However, this does not mean that the hardware of intelligent machines can never be coupled with the human brain. Instead, it shows that strengthening the hardware connection between the human brain and intelligent machines is an attractive direction for intelligent development.

The intelligence of AI lies in its software. Why is AI possible? How is it related to human brain intelligence? To answer these questions, we need to carry out software analysis. Thinking is the function of the human brain. The analysis of thinking can be carried out through the functional structure method, hierarchical method, or the two approaches combined. From the hierarchy perspective, thinking has several layers: the first one is the layer of form, which is language; and the second one is the layer of content, which is the consciousness concept. These two layers are inward analysis of thinking itself. From the perspective of its interaction with its applied object, there is another layer of thinking, i.e., its outward function layer. When these three layers are divided according to a functional structure, language and consciousness concept are the internal structures, while the functional layer boasts the external function of thinking and consciousness, namely, intelligence. Consciousness concept, as the internal structure of thinking, refers to the structure of consciousness. Depending on the internal structure, reflection and control activities have corresponding manifestations at all internal levels of thinking. In the content level, it is shown as the activities of various consciousnesses and in the form level as corresponding language activities. Without the activities at all internal layers of thinking, there would be no overall intelligence or activity. Activities at the language layer are coding activities inside the human brain, which can be carried out directly by the bioelectrical movement. Therefore, when people think of specific words, there are always corresponding specific electrical signals in the brain.

The fact that thinking activities have specific manifestations at the language level shows that thinking can be formalized through language. Language is not a conscious or conceptual thing. Instead, it has a corresponding movement of electrical signals in the brain, showing the characteristic of physical movement. As far as language and electrical signals are concerned, language is the content while electrical signals are the form. Language outside the brain refers to the sound that can be heard and words that can be seen. The internalization of language outside the brain is completed by the brain's thinking language, whose existence depends on the physiological, chemical, and physical movements of the brain. When dissecting a human brain, there is no language entity inside. Instead, only corresponding brain structures and movement can be found. Therefore, the role of the language inside the brain is highlighted as the coding rules of electronic movement. Only when there is a specific movement of electrical signals in the brain can there be language movement and further consciousness activities. These serial relationships lay a foundation for formalizing thinking activities, namely, linguisticization, and then for the electronic movement. This is the most fundamental reason why electronic computers are able to simulate the intelligence of the human brain.

In short, AI is not independent. It is within the scope of human intelligence and part of it. The reason why AI belongs to human intelligence is that it is a product of human intelligence developed to a certain historical stage, a tool of the human brain, and an expansion of human intelligence [5]. All AI is extending along the direction of human intelligence, and all their functions in various fields of social life are within the scope of reflection and control. With the development of humankind, all aspects of social life today are increasingly complex, which makes it more and more difficult for the human brain to make direct adjustment and control and increasingly unable for mankind to meet the requirements in both magnitude and precision. As a result, the activities in many fields of society cannot be regulated and controlled only by human intelligence. Of course, this is not the limitation of the human brain in quality, but in quantity. Therefore, the development of human intelligence can be achieved by brain expansion in a certain way, so as to meet the requirements of social life.

3. AI and Human Consciousness

In the coming years, machines will get smarter. If we cannot distinguish a machine from a human, then we have reason to think that this machine is intelligent [6]. Therefore, the question we are going to face is as follows: can an intelligent machine be considered as having consciousness? This requires us to understand the relationship between AI and human consciousness.

3.1. AI Promotes the Development of Human Intelligence

As a necessary supplement to human intelligence, AI effectively extends the human brain and enlarges its intelligence. AI and the human brain are correlated and have been supporting each other forward. All these make human cognition scope continuously expand to the micro- and macropoles, enabling people to have a deeper understanding of the essence of things indirectly, and greatly enriching the content of consciousness.

AI, which simulates human operational intelligence, is far superior to human beings in computing speed, capacity, and accuracy. It indeed can liberate mental labor. With the support of the Internet and big data technology, AI will help humans in more fields and more profoundly, even conducting rescue operations in extreme environments. In the field of medical practice, brain stimulation is helpful to restore the damaged brain nerve. In terms of transportation, with the application of data, connection, real-time sensing, and traffic prediction, humans will experience shared riding and automatic driving for the first time. The revolution of the third-generation culture carrier represented by AI will promote the great change of human memory and learning style. AI is our brain assist device, which stores a lot of information in an intact manner. The undertaker of memory and thinking is gradually separating from the human body and tends to become objectified. Portable computers have replicated what we call as cognition, and even our human rationality also faces challenges. However, human beings can use their dynamic intuition and give play to their innovation abilities [5]. The printing carrier once ushered in a period of flourishing human culture. Now, we should keep an open mind on the promotion role of AI in human consciousness. AI technology has put forward new requirements for human data observing and processing ability. The workers who receive information technology (IT) vocational training will better adapt to the changes, so we can be one step forward in the transformation to an intelligence-intensive society. With the help of intelligent machines, humans could become new-type, creative, and reliable cognitive subjects.

3.2. Human Consciousness Restricts the Development of AI

The nature of consciousness affects the development of AI. Consciousness can be introspective, which reveals what cannot be reached by the objective research on consciousness. Human consciousness is not a passive or negative reflection of reality; instead, it is a positive and active one. When determining the behavior of the subject, the external experience must be reflected through the inner world as well as the thinking and feeling system of the subject. The so-called animal consciousness is a statement untested, because animals cannot distinguish themselves from their activities. They are integral. The same is true of artificial consciousness. Although AI can complete part of human thinking activities, it does not understand the meaning of doing this. It operates mechanically and aimlessly. Even if AI has a purpose, it is all instilled by humans to achieve goals of themselves. After 70 years, the movement of logical functionalism ended in a dismal way, while the structuralism of consciousness points out a new direction for AI. Structuralism has a successively experienced semantic network and neural network. The latter one argues that connections among things in the world are all the same, while the differences lie in their frequencies of occurrence. The neural network cannot distinguish the “White” as a name and the “White” as a color. This kind of AI, in essence, is a program or function that makes similar reflex responses to specific stimuli. AlphaGo is weak AI, and programming is not an effective way to achieve a machine’s consciousness.

The change and development of consciousness bring about corresponding changes in AI. Under the consciousness theory system of a subject-object dichotomy, abstract operational rules promote the R&D of algorithms. This theory argues that human consciousness can be

simply summarized as the brain's symbolic operation, while the characteristics of consciousness in intuition, common sense, and external environment are ignored. In fact, the change in AI development level is indeed related to the development of human consciousness. Behaviorism and structuralism, which are similar to the human brain neural network, are simulations of human adaptive mechanisms and not restricted by forms. Subject-object integration, a philosophical trend focusing on the interaction with the real world, provides ideological inspiration for AI. Human consciousness theory criticizes principles of form. Deep learning represented by AlphaGo has gotten rid of structural restrictions and acquired problem-solving strategies through learning human's experiences. In 2016, by adopting Monte-Carlo Tree Search (MCTS) and deep learning (DL) to think from the whole picture and make the optimal choice, AlphaGo, a program developed by DeepMind, won the Go game, getting rid of the restrictions of brute force method [7]. When embodiment philosophy arose, AI began to imitate the human body's movement and gestures, such as simulating the rules of facial movement. To some extent, the fact that AI began to shift its focus to the human body and external environment is inseparable from the exposition of consciousness by phenomenologists [8]. We can tell that although the philosophical theory of consciousness cannot directly improve the technical essence of AI, its development and change will provide foresight for the exploration of AI.

4. AI and Cognitive Computing Technology

AI is a broad concept. From the perspective of ultimate goals, cognitive computing is an important way to realize AI. Cognitive computing refers to cognition and effective expression of the internal meaning of the objective world as well as of various pieces of information and data that can be observed and measured at present. It is the expression of AI toward specific problems to be solved.

4.1. The Concept of Cognitive Computing

Cognitive computing is a technique that enables humans to cooperate with machines. This term comes from cognitive science and artificial intelligence. It builds algorithms with theories of cognitive science to simulate human's objective cognition and psychological cognition process, so as to enable machines to reach a certain degree of "brain-like" cognitive intelligence [9]. Cognitive computing uses technology and algorithm to automatically extract concepts and relationships from data, understand their meanings, learn independently from data patterns and prior experience, and ultimately extend what people or machines could do on their own. Based on this, Roma further put forward three main applications of cognitive computing: robotic and cognitive automation to automate repeatable tasks to improve efficiency, quality, and accuracy; cognitive insights to uncover hidden patterns and relationships to identify new opportunities for innovation; and cognitive engagement to drive customer actions by delivering hyper-personalization at scale [10]. Cognitive computing is a synthesis of technologies where each of them contributes a distinct methodology for addressing problems in its domain. Artificial Neural Network (ANN) uses the interactions of biological neurons as a model for pattern recognition, decision, modeling, and forecasting. Fuzzy logic uses approximate

information in a manner similar to the human decision process and is useful in control and decision-making applications. Evolutionary computation adopts natural selection and evolution theory and is useful in optimization. Cognitive computing provides an effective way to analyze technological processes and human activities [11].

Based on the above concepts, cognitive computing can be simply understood as a technical field that integrates multiple technologies and aims to use artificial mechanisms based on computing technology to realize the human cognitive function. It is the core technical field of cognitive science. In essence, cognitive computing is expected to understand the internal relationships among various kinds of data and phenomena in the real world through technologies, such as AI, pattern recognition, and machine learning, and further develop tools and systems to improve productivity, protect the environment, and contribute to social governance.

4.2. Characteristics and Advantages of Cognitive Computing

After tabular computing and programming computing, now the era of cognitive computing is coming. Generally speaking, cognitive computing has a wide range of applications, including participation, decision-making, discovery, and other aspects, centering the improvement of human's "cognitive" ability. Leslie G. Valiant from Harvard University thought that compared with other approaches, cognitive computing has three main characters: each act of memorization, learning, or recall is an algorithmically simple process executed on a network laden with information previously acquired; the system learns continuously as a background activity; in more complex cognitive processes such as analyzing complex scenes or reasoning, the internal computations have an important time domain and state information needs to be retained [12]. The cognitive computing system has a strong comprehension ability. Through natural language comprehension technology and its superior ability to process structured and unstructured data, it can interact with users in various industries and then understand and respond to their problems. The cognitive computing system has an intelligent logical thinking ability. It can reveal insights, patterns, and relations through data and hypothesis and connect scattered pieces of knowledge for reasoning, analysis, comparison, induction, summary, demonstration, and obtainment of deep insights and evidence for decision-making. The cognitive computing system has excellent learning ability. Through evidence-based learning ability, it can rapidly extract key information from big data and learn like a human. It can gain feedback through expert training and experience learning in the interaction to optimize models and make improvements. In addition, a cognitive computing system also has elaborated personalized analysis ability. Using text analysis and psycholinguistic models, it can conduct an in-depth analysis of massive social media data and business data, grasp users' personalities, and portray individuals in an all-around way. This system is not a simple collection of all these technologies. Instead, it integrates these technologies in an unprecedented way, profoundly changing the means and efficiency of solving business problems.

Compared with previous computing paradigms, cognitive computing has significant characters in adaptability, interaction, iteration, and context sensing. It can perceive the surrounding

environment and context and make the corresponding self-adaptation. Cognitive computing requires dynamic programming and must understand, identify, and extract context elements, such as connotation, grammar, time, location, regulation, user profile, process, tasks, and targets. They might use multiple information sources, including structured and unstructured digital information, and sensorial inputs, such as vision, gesture, hearing, or information of the sensors. Cognitive computing also has the “memory” function and is able to conduct an iterative operation. The cognitive computing system must be able to remember previous interactive information to make rational reasoning and aid decision-making through the superposition of information and semantics. For instance, as a digital medical aid, when a user communicates with it about the personal situation of “chest distress or insomnia” at 1:00 a.m., the medical aid must “recognize” the current time and the user’s situation, make comprehensive judgments combining user’s previous conditions, and offer a reasonable suggestion.

At present, there are four key technologies of cognitive computing recognized by researchers: first, machine learning, natural language understanding, and human-computer interaction techniques which are on the top floor; second, big data technologies, including how to store, organize, manage, and analyze big data; third, computer architecture (the computing ability required by the cognitive system is far more than what we can provide today; therefore, how to realize the design of a data-centered system is also a challenge facing us today); fourth, at the bottom level breakthroughs of atomic and nanotechnologies which are required [13]. There are two main tasks of cognitive computing: one is to study and simulate a human’s understanding of the objective world through a computer; the other is to take the cognition and value discovery of information and data as the main goal. Compared with AI, the research of cognitive computing is deeper and more specific. “Deeper” means that it does not only study the simulation of human brain behavior but also focus on the understanding of the operation law of the objective world as well as the internal law and external expression of the data generated in the world; “more specific” means that it has more direct expressions in the applied business area, which can offer direct decision-making suggestions to corporate leaders.

5. Forecast of Future Development of AI

What kind of role will AI play in the future? Professor Jiang from Stanford University said, “I hope young people can understand the difference between the working mode of AI and that of humans, and then develop the abilities that can distinguish them from AI.” [14]. From the perspective of core technologies, breakthroughs in three levels are expected to advance further development of AI. The breakthroughs in platforms, algorithms, and interfaces will promote AI to achieve leapfrog development. Building an intelligent platform that can serve various enterprises and meet different demands will be a major trend of future technological development.

5.1. Making Innovations of Intelligent Technologies and Realizing Reproduction of Human’s Personalized Intelligence

At present, AI is still in the “weak AI” phase and can only simulate, extend, and expand low-end human intelligence, namely, human’s feelings, perceptions, and conventional and programmed logical reasoning. As for human’s high-end creative intelligence, imagination, intuition, potential, and unconventional and nonprogrammed personalized intelligence such as tacit knowledge, experience, and skills which can only be expressed through behavior, “weak AI” is not able to simulate them, let alone extension or expansion. This high-end and unconventional intelligence can only be simulated, extended, and expanded when we enter into a “strong AI” phase.

The development of AI is, in varying degrees, based on the research of brain science. The inexplicable “black box” characteristics of its operation are closely related to the fact that brain science has not yet fully grasped the operation rules and mechanisms of human brain intelligence. The cracking of AI’s “black box” characteristic depends on the further development of brain science. Will AI surpass man’s biological intelligence in the future? This kind of judgment actually presupposes that machine intelligence has great development potential, but it ignores the fact that brain science has already revealed that human brain intelligence is also far from being fully exploited and released and also enjoys great potential. In terms of AI, the proportion of human intelligence that can be simulated or extended is just the tip of the iceberg above the water, only including the conventional, logical, explicit, and universal consciousness and intelligence. The vast number of unconventional, illogical, and personalized consciousnesses hidden in the water is still difficult to be simulated or extended. As is pointed out by the “Iceberg Theory” of the psychoanalyst Freud, the human psychological structure is composed of consciousness, preconsciousness, and subconsciousness. Consciousness is only the tip of the iceberg above the water, while subconsciousness occupies most of the psychological structure and hides below the water surface [15]. This is mainly because humans have little cognition and development of this part of consciousness and intelligence, not to mention the artificial intelligence of it. Of course, with the continuous development of the brain, human potential, and AI itself, it is possible for AI to simulate and extend human’s personalized and fuzzy consciousness and intelligence. American futurist Ray Kurzweil once predicted that the year 2045 will be a time of profound and divisive transformation, “abiotic intelligence in this year will be a billion times the wisdom of all human beings today” [16].

How to realize strong AI? This requires a “brain-computer interface” (BCI), a cutting-edge research area. It studies the means of establishing a direct connection between human or animal brains and external devices to translate consciousness in real time and ultimately transfer and download thoughts freely among humans or between humans and machines. Astonishingly, Neuralink, a brain-computer interface research company founded by Elon Musk, released a groundbreaking “brain-machine interface” technology, which utilizes threads 4 to 6 μm thick, less than a tenth of that of hair, to transmit brain signals fetched by chips. Tiny electronic devices are implanted in the brain so that thoughts can be transmitted through wireless devices and even interact with iPhone apps. Brain-computer interface technology can realize four functions. By manipulating machines through the mind, machines can replace some functions of the human body and repair the physical defects of the persons with disabilities. Anything can be controlled through the mind. Brain operation can be improved through BCIs, making us feel like we have just had a good sleep and feel energized, focused, and quick on the trigger so that we can work soberly and efficiently. In 2014, ABM, an American

company, trained testers through an EEG brain-computer interface, making novices learn 2.3 times faster than before. Through the brain-computer interface, we can acquire a lot of knowledge and skills in a short time and even acquire superpowers that ordinary humans cannot possess. With a brain-computer interface, human beings can communicate with each other without using language and only rely on neural signals in the brain, thus realizing “lossless” brain information transmission. Musk crazily envisions that human-computer combination can help realize faster and more accurate communication, and this nonverbal communication is better. However, we believe that communication is a basic human behavior and also the basis of human cooperation. In a virtual digital space, the human language will evolve rather than be replaced entirely. The future development of AI is to enhance, not to replace, the overall intelligence of human beings and promote the complementation of AI and human intelligence, giving play to their respective advantages to realize the “coevolution” of human and AI machines.

5.2. Making Breakthroughs in Specialized Algorithms and Building Intelligence Service Space

AI is expected to soon obtain an almost infinite information storage space, quantum computing ability 100 times or even 10,000 times more than that of humans, and breakthroughs in various specialized algorithms. Even if an AI system is equipped with the most advanced current computing platform, without effective algorithms, it will only be like a person with well-developed limbs but the head of a moron, which cannot be regarded as having real intelligence. The enhancement of algorithm capability will further promote the continuous breakthroughs of AI. In many consumption scenarios in life, people’s demand for personalized experience is increasing, and personalized and scene-based services will gradually become the main direction of AI-driven innovation. With the help of the Internet, a professional-level knowledge base together with program settings is expected to “answer” most professional or scientific questions in the near future. Their professional abilities are equivalent to those of senior doctors, architects, engineers, math professors, and so on.

In the field of education, AI can become an important driver for the education of students. It has unique advantages in creating large-scale personalized learning environments and building smart campuses. With the deep integration of education with cloud computing, big data, VR/AR, and other technologies, the application of AI in the field of education has infinite potential and possibility in the future [17]. At present, BCI technology is applied to measure students’ implicit data, including their learning status, attention level, cognitive load, and learning style [18, 19]. Brain measurement can reveal the brain differences among students and enhance our understanding of learning [20]. Based on the attention data collected when different learners watched, listened, read, wrote, or operated learning materials, the SVM learning algorithm was used to identify their learning styles. Its average recognition rate was 75.8%, and the highest accuracy of a single time was 83.3% [21]. In the future, cognitive computing will be mainly used to customize personalized learning assistants for students to improve their overall learning experience. Through collecting and analyzing students’ learning data, AI gradually outlines each student’s learning style and characteristics and then automatically adjusts teaching content,

method, and pace, so that students can have access to the most suitable education for themselves. Not only does an intelligent learning system easily provide personalized teaching for students, but also real-time data and automatic analysis provide teachers with a lot of information so that they can have a deeper understanding of each student. Meanwhile, data can guide teachers to constantly improve their teaching content and make more accurate teaching planning based on the problems of each student, so that their teaching could be more targeted. However, it should be noted that everything will become different in the future. Only the ability to solve new problems remains unchanged [21]. 98 percent of our genetic make-up is the same as that of chimpanzees, but our language, values, artistic expression, understanding of science, and research of technologies make us distinctive. This is the result of creativity, which is credited with most of the interesting, important, and humanized things. Creative people tend to be more independent in thinking and operations. They tend to be imaginative, curious, and willing to try and take risks. In its competition with the Go genius Jie Ke, AlphaGo was like a martial arts master who could absorb the opponent's power. Each manual and game could be an effective source driving its power growth. It was so powerful that even one of the smartest humans felt "cold." Moreover, robots never get angry or tired, so it is impossible for humans to win the competition with machines. Therefore, "we cannot just instill our children with knowledge, because machines can learn faster [22]."

In the health care area, many companies are developing AI technologies that can be used in the medical field. Classification and identification capabilities are the primary goals of these devices. It can be very helpful in identifying carcinogens because it can help inspectors interpret countless documents in less time. In addition, it can evaluate relevant information of patients, go through all the medical records thoroughly, find clues that may cause the problems of patients, and assist diagnosis. Through the analysis of the medical history of patients based on lots of medical textbooks and related data, AI can provide some diagnostic bases for doctors. Sometimes, they may offer disease diagnosis that doctors have never considered or even do not know. In the business area, combined with big data and algorithms which can meet the demands of customers, AI can play an important role in the economic decision-making process. Companies can use AI to incorporate various risk factors into their decision-making process and then offer effective suggestions on investment or site selection of branches. In the field of finance, AI promotes the rise of humanized fin-tech, which can be widely adopted by banks and financial institutions to serve growing numbers of customers more effectively. Besides automatically performing back-end and administrative tasks, it can also take the initiative in activities toward customers. In the manufacturing industry, AI can assist designers to complete the design of products. Ideally, it can largely make up for the shortage of medium and high-end designers, thus greatly improving the product design capabilities of the industry. Meanwhile, by mining and studying a large amount of production and supply chain data, AI is expected to help optimize the allocation of resources and improve the efficiency of enterprises. In an ideal situation, AI can provide enterprises with whole-process support including product design, raw material procurement plan and distribution, production and manufacturing, and user's feedback data collection and analysis, so as to promote the transformation and upgrading of China's manufacturing industry.

5.3. Demonstrating Value Rationality and Avoiding Ethics Risks

Understanding the trend of technological and ethical risks of AI is a part of the research on the development trend of AI. And how to effectively avoid these risks is also a very popular research topic in the field of AI.

In various AI application scenarios such as data analysis, content recommendation, and face recognition, human identity and behavior are directly involved and affected. The harm and negative impact of abusing related technologies will be far greater than that of traditional networks and digital technologies [23]. Specifically, there is algorithmic bias in AI applications. An algorithm is essentially an objective mathematical expression, but modeling and data input are completed by humans. In this process, human's inherent bias and discrimination will affect the decision-making of the algorithm and then lead to ethical problems such as algorithmic bias. There are social ethical problems with AI. The application of AI in human life makes the relationship between humans and machines increasingly complicated, further leading to a series of social-ethical problems. Among them, the most primary one is whether AI can replace humans to work? From the perspective of operational effectiveness and economic efficiency, a large number of white-collar or blue-collar jobs will be occupied by AI in the future, which will inevitably lead to large-scale technical unemployment and, to a certain extent, lead to the antagonism between human and AI. There are responsibilities and emotional ethical problems with AI. The widespread application of AI brings about the difficulty in judging subjects in many events or cases. For example, during the sudden braking of driverless automobiles in case of emergency, each step adopted by AI is stipulated by algorithms, which are set by human beings. Therefore, ethical problems are bound to arise in the judgment of liability subjects in scenarios like this. If we stick to the position of optimism and pragmatism without introspection, the potential threats of AI to society and individuals will be overlooked or ignored and the ethical risks will not be prevented in advance or corrected afterward. It will inevitably not only lead to serious consequences but also destroy the trust of the whole society in AI, which will inevitably bring on huge risks and losses to the institutions and enterprises within the innovation ecosystem.

In order to standardize the development of AI, the UK Government Office for Science published the report "Artificial Intelligence: opportunities and Implications for the Future of Decision Making" in 2015, which pointed out the serious consequences of algorithmic bias, nontransparency, and improper accountability and stressed that further development of AI should be based on the premise of enabling innovation, building trust among citizens, establishing a stable environment, and fostering appropriate access to the data necessary [24]. In December 2016, the Institute of Electrical and Electronics Engineers released "Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems, First Edition," which suggested that "human benefit," "responsibility," "transparency," and "education and awareness" should be the general principles in the research and development process of products [25]. In June 2019, the "G20 AI principles" were issued in the communique of the G20 Ministerial Meeting on Trade and Digital Economy, which made it

clear that AI systems should be “stable and secure” in the whole life cycle and put forward five principles for responsible stewardship of trustworthy AI [26]. In the same month, the “Governance Principles for the New Generation Artificial Intelligence—Developing Responsible Artificial Intelligence” was released by the National Governance Committee for the New Generation Artificial Intelligence of China, proposing eight principles of “harmony and human-friendly,” “fairness and justice,” “inclusion and sharing,” “respect for privacy,” “safety and controllability,” “shared responsibility,” “openness and collaboration,” and “agile governance” [27].

Most of the above documents consider the possible impact of AI technology applications on society from the perspective of national development strategy. Science is a powerful instrument. How it is used, whether it is a blessing or a curse to mankind, depends on mankind and not on the instrument [28]. Generally speaking, the ethical norm of AI aims to equip AI with a “good core (conscientiousness).” This means that, different from other technologies, the ethical research of AI should be carried out centering “machine’s core” and “human conscientiousness.” “Machine’s core” mainly refers to the moral algorithm of AI, which aims to instill “good ethics” into AI so as to generate moral AI or machines. “Human conscientiousness” mainly refers to the design and application ethics of AI, which aims to enable the developers and users of AI to have “conscientiousness,” ensuring the design of AI conform to morality, avoiding malicious design, and making sure that AI is used rationally to benefit human society. Whether AI is a “blessing” or a “curse” to mankind depends on what kind of value human holds. Only by highlighting the value rationality in the design of AI and avoiding the distortion of its development, we can realize the sustainable development of human beings and finally achieve human liberation.

6. Conclusion

Based on the above analysis, the following conclusions are drawn:

- (1)

In the future, the human-computer combination can help realize faster and more accurate communication, and this nonverbal communication is better. However, we believe that communication is a basic human behavior and also the basis of human cooperation. In a virtual digital space, the human language will evolve rather than be replaced entirely. The future development of AI is to enhance, not to replace, the overall intelligence of human beings and promote the complementation of AI and human intelligence, giving play to their respective advantages to realize the “coevolution” of human and AI machines. Breakthroughs in algorithms represented by cognitive computing promote the continuous penetration of AI into fields such as education, commerce, and medical treatment to build up AI service space. As to human’s concern, namely, who controls whom between humankind and intelligent machines, the answer is that AI can only become a service provider for human beings, demonstrating the value rationality of following ethics

- (2)

If we view the emergence of AI from the historical process of human inventing, manufacturing, and using tools, we will find that AI is an intelligent tool. Though it has qualitative differences from previous physical tools, based on their nature as tools, they have some same characteristics. In the competition between man and AI, AlphaGo's victory, in fact, is not the victory of intelligent robots but the victory of many present and past Go experts. Therefore, we cannot simply draw the conclusion that the intelligence of robots is higher than that of humans or that robots can replace or dominate humans. We should view AI from the history of human inventing and using tools and from the relationship between humans and tools. From the perspective of tool theory, the earliest tools of mankind are stone implements made by the primitive ancestors. After entering the agricultural civilization, farmers invented, manufactured, and used various farm tools. After entering the industrial civilization, engineers invented and manufactured a variety of automated machines. Stone implement, farm tools, and machines are all the replacement, extension, and expansion of human physical ability. Every qualitative leap of physical tools is a huge liberation of human physical fitness, improving productivity and promoting social development and progress. In the postindustrial civilization era, most of the original artificial physical fitness has been transformed and replaced by artificial intelligence. The birth of AI is another revolution in human tools' history. AI is not only automated but also intelligent. It can replace, extend, and expand not only most of the human physical fitness but also part of human intelligence. It frees humans from not only burdensome and tedious physical labor but also part of the mental labor, once again rapidly improving productivity and promoting human development and social progress

- (3)

Intellectualization is the inexorable trend of the future whether seen from the maturity of the technology or seen from the development of human intelligence. As Kurzweil said, we could neither prevent the acceleration of change nor prevent AI from surpassing humans in various fields [17]. The most unique feature of AI technology is that it can endow "machine" with "intelligence." Technologies were invented by "human intelligence" before. However, from now on, technologies can also be invented by "the machine's intelligence." An important mark of the coming of AI singularity is "producing intelligence with machine's intelligence" [29]. By 2050, AI will be infinitely close to human intelligence. However, mankind "still has the right to determine future technologies and life." In other words, many advanced abilities of human beings are irreplaceable by AI. AI can only be a service provider for human beings. Although from the fractional and individual level, AI can replace, expand, and surpass humans; its position as a subject is stronger than that of a human, namely, it can dominate and control individuals, and humans serve AI. However, from the perspective of the whole and mankind, AI is an electronic machine, not an animate individual, having no life form or movement. It has no independent needs, attributes, nature, consciousness, or social behavior like human beings. Therefore, it is impossible for AI to become a subject like a human. On the contrary, it can only be a machine used by people, being tools and accessories in human production and life. If we excessively rely on AI, we will have a

dependence tendency, and this tendency (for example, we think that the data provided by advanced instruments are the most perfect and reliable) will not only make people lose their critical thinking skills but also cause irreversible mistakes or disasters. Therefore, for human beings and society, many key abilities, such as the ability to promote social and human progress, can only rely on ourselves

In human history, all major technological revolutions have brought about shocks to humankind and even the whole society. The invention of ironware and the beginning of traditional agricultural society sparked a war in the cold weapon era. In China, this war lasted from the Spring and Autumn Period and the Warring States Period to the Han Dynasty. The invention of electricity and the arising of industrial society incurred World War I and World War II, staining the first half of the 20th century with blood. Now, while delivering technological benefits, the advancement of technologies such as AI, nanotechnology, brain-computer interface, and biotechnology will also start a new storm of social change. As what Charles Dickens said in *A Tale of Two Cities*, "it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way" [\[30\]](#).

- [Latest news](#)
- [Unread news](#)
- [Subscribe](#)

Science X Account

Remember me

Sign In

[Click here to sign in with or](#)

[Forget Password?](#)

[Not a member? Sign up](#)

[Learn more](#)

- [Nanotechnology](#)
- [Physics](#)
- [Earth](#)
- [Astronomy & Space](#)
- [Chemistry](#)
- [Biology](#)
- [Other Sciences](#)

- [Medicine](#)

- [Technology](#)

It appears that you are currently using **Ad Blocking software**. What are the consequences?×

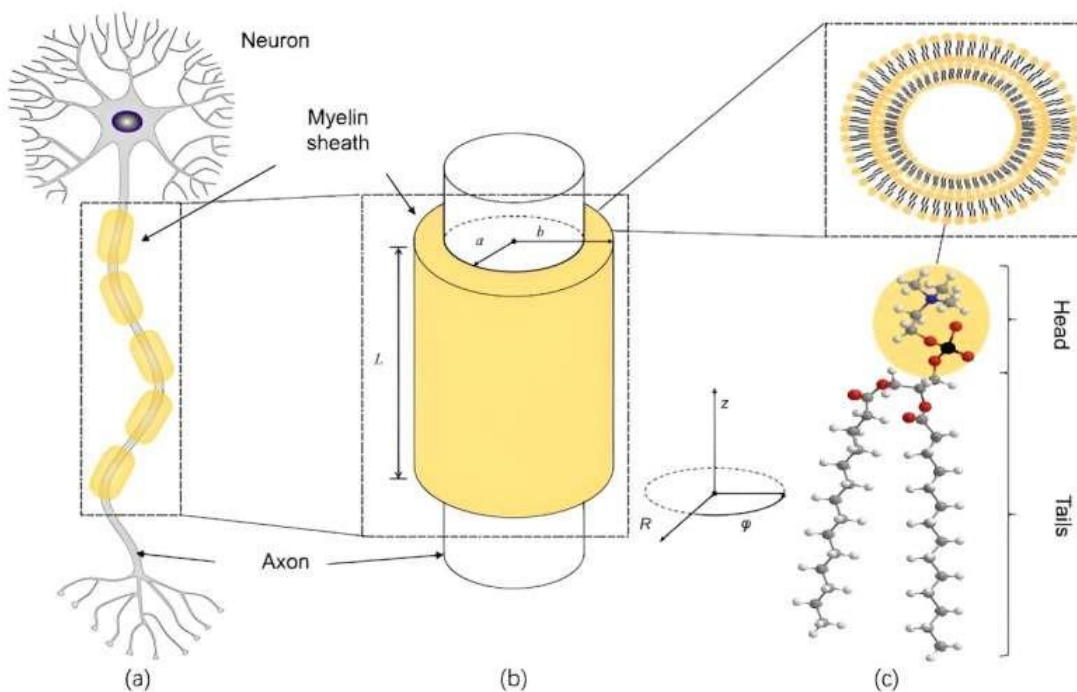
- [1.4K](#)
- [Twit](#)
- [Share](#)
- [Email](#)

1. [Home](#)
2. [Physics](#)
3. [General Physics](#)
1. [Home](#)
2. [Physics](#)
3. [Quantum Physics](#)

-
-

Photon entanglement could explain the rapid brain signals behind consciousness

by David Appell , Phys.org



A

schematic showing, left to right (a) a neuron with multiple myelin sheaths along its length, (b) modeling of a myelin sheath segment encasing an axon segment, and (c) phospholipid molecules, a major component of myelin, with a tail consisting of a large number of carbon-hydrogen (C-H) bonds. Credit: *Physical Review E* (2024). DOI: 10.1103/PhysRevE.110.024402 Understanding the nature of consciousness is one of the hardest problems in science. Some scientists [have suggested](#) that quantum mechanics, and in particular quantum entanglement, is the key to unraveling the phenomenon.

Now, a research group in China has shown that many [entangled photons](#) can be generated inside the myelin sheath that covers nerve fibers. It could explain the rapid communication between neurons, which so far has been thought to be below the [speed of sound](#), too slow to explain how the neural synchronization occurs.

The paper is [published](#) in the journal *Physical Review E*.

"If the power of evolution was looking for handy action over a distance, quantum entanglement would be [an] ideal candidate for this role," said Yong-Cong Chen in a statement to Phys.org. Chen is a professor at the Shanghai Center for Quantitative Life Sciences and Physics Department at Shanghai University.

The brain communicates within itself by firing [electrical signals](#) called synapses between neurons, which are the main components of nervous tissue. It is the synchronized activity of millions of neurons that consciousness (among other brain business) relies on. But the way this precise synchronization takes place is unknown.

Connections between neurons are called axons—long structures akin to electrical wires—and covering them is a coating ("sheath") made of myelin, a white tissue made of lipids.

Comprised of up to hundreds of layers, myelin insulates the axons, as well as shaping them and delivering energy to the axons. (In actuality, a series of such sheaths stretches across the length of the axon. The myelin sheath is typically about 100 microns long, with 1 to 2 micron gaps between them.) [Recent evidence suggests](#) myelin also plays an important role in promoting synchronization between neurons.

But the speed at which signals propagate along the axons is below the speed of sound, sometimes much below—too slow to create the millions of neuron synchronizations that are the basis for all the amazing things the brain can do.

To remedy this problem, Chen and his colleagues investigated if there could be entangled photons within this axon-myelin system that could, though the magic of quantum entanglement, communicate instantly across the involved distances.

A tricarboxylic acid cycle releases energy stored in nutrients, with a cascade of infrared photons released during the cycling process. These photons couple to vibrations from carbon-hydrogen (C-H) bonds in lipid molecules and excite them to a higher vibrational energy state. As the bond then transitions to a lower vibrational energy state, it releases a cascade of photons.

The Chinese group applied cavity quantum electrodynamics to a perfect cylinder surrounded by the myelin, making the reasonable assumption that the outer wall of the [myelin sheath](#) is a perfectly cylindrical conducting wall.

Discover the latest in science, tech, and space with over **100,000 subscribers** who rely on Phys.org for daily insights. Sign up for our [free newsletter](#) and get updates on breakthroughs, innovations, and research that matter—**daily or weekly**.

Subscribe

Using quantum mechanical techniques, they quantized the [electromagnetic fields](#) and the electric field inside the cavity, as well as the photons—that is, treated them all as quantum objects—and then, with some simplifying assumptions, solved the resulting equations.

Doing so gave the wavefunction for the system of the two photons interacting with the matter inside the cavity. They then calculated the photons' degree of entanglement by determining its

quantum entropy, a measure of disorder, using an extension of classical entropy developed by the science polymath John von Neumann.

"We showed that the two photons can indeed have higher rate of being entangled under occasions," said Chen in his statement.

The conducting wall limits the electromagnetic wave modes that can exist inside the cylinder, making the cylinder an electromagnetic cavity that keeps most of its energy within it. These modes are different from the continuous electromagnetic waves ("light") that exist in free space.

It is these discrete modes that result in the frequent production of highly entangled photons within the myelin cavity, whose rate of production can be significantly enhanced compared to two untangled photons.

Entanglement means the two-photon state is not a classical combination of two photon states. Instead, measuring or interacting with one of the photons instantly affects the same property of the second photon, no matter how far away it is.

Entanglement has been demonstrated for a system whose members are [over 1,000 km apart](#). Nothing like it exists in classical physics; it is purely a quantum phenomenon. Here entanglement would raise the possibility of much faster signaling along the sections of myelin that encase segments of the axon's length.

One possibility, the authors write, is that the entanglement of photons could transform into entanglement along potassium ion channels in the neuron. If so, the opening and closing of one channel may affect the performance of another somewhere else.

Chen told Phys.org that their result is a combination of two phenomena that exist but are still largely mysterious: consciousness (let alone quantum consciousness) and [quantum entanglement](#).

"We won't say there is a direct connection. At this early stage, our primary goal is to identify possible mechanisms of neural synchronization, which affects numerous neurobiological processes. Through this work, we hope to gain a better understanding."

More information: Zefei Liu et al, Entangled biphoton generation in the myelin sheath, *Physical Review E* (2024). DOI: [10.1103/PhysRevE.110.024402](https://doi.org/10.1103/PhysRevE.110.024402). On arXiv: DOI: [10.48550/arxiv.2401.11682](https://arxiv.org/abs/10.48550/arxiv.2401.11682)

Journal information: [Physical Review E](#) , [arXiv](#) © 2024 Science X Network

A New Approach to the Hard Problem of Consciousness: a Quasicrystalline Language of “Primitive Units of Consciousness” in Quantized Spacetime

- Published: 10 July 2020
- Volume 62, pages 48–68, (2020)

[Klee Irwin](#)

Abstract

The hard problem of consciousness must be approached through the ontological lens of twentieth-century physics, which tells us that reality is information theoretic Wheeler (*American Scientist*, 74, 366–375, [1986](#); Wheeler ([1990](#)) and quantized at the level of Planck scale spacetime Snyder (*Physical Review D*, 67, 38–41, [1947](#)). Through careful deduction, it becomes clear that information cannot exist without consciousness—the awareness of things. And to be aware is to hold the meaning of relationships of objects within consciousness—perceiving abstract objects, while enjoying degrees of freedom within the structuring of those relationships. This defines consciousness as language—(1) a set of objects and (2) an ordering scheme with (3) degrees of freedom used for (4) expressing meaning. And since even information at the Planck scale cannot exist without consciousness, we propose an entity called a “primitive unit of consciousness,” which acts as a mathematical operator in a quantized spacetime language. Quasicrystal mathematics based on E8 geometry Sadoc and Mosseri (*Journal of Physics A: Mathematical and General*, 26, 1789, [1993](#)) seems to be a candidate for the language of reality, possessing several qualities corresponding to recent physical discoveries and various physically realistic unification models.

This is a preview of subscription content, [log in via an institution](#) to check access.

References

- Alfonso-Faus, A. (2011). Cosmology, holography, the brain and the quantum vacuum. *arXiv preprint, arXiv*, 1102.0443.

[Google Scholar](#)

- Bem, D. J.. (2011) "Experimental evidence for anomalous retroactive influences on human cognition and affect." QUANTUM RETROCAUSATION: THEORY AND EXPERIMENT. Vol. 1408. No. 1. AIP Publishing.
- Bem, D., et al. (2014) "Feeling the Future: A Meta-Analysis of 90 Experiments on the Anomalous Anticipation of Random Future Events." Available at SSRN 2423692 .
- Bohm, D., & Hiley, B. (1977). *On the intuitive understanding of non-locality as implied by quantum theory*. Netherlands: Springer.

[Book Google Scholar](#)

- Calcagni, G. (2010a). Fractal universe and quantum gravity. *Physical Review Letters*, 104(25), 251301.

[Article PubMed Google Scholar](#)

- Calcagni, G. (2010b). Quantum field theory, gravity and cosmology in a fractal universe. *Journal of High Energy Physics*, 2010(3), 1–38.

[Article Google Scholar](#)

- Castro, C.. (2002). "FRACTAL STRINGS AS THE BASIS OF CANTORIAN-FRACTAL SPACETIME AND THE FINE STRUCTURE CONSTANT." arXiv preprint hep-th/0203086.
- Castro, C. (2007). A CHERN–SIMONS E8 GAUGE THEORY OF GRAVITY IN D=15, GRAND UNIFICATION AND GENERALIZED GRAVITY IN CLIFFORD SPACES. *International Journal of Geometric Methods in Modern Physics*, 4(08), 1239–1257.

[Article Google Scholar](#)

- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.

[Google Scholar](#)

- Chamseddine, A. H., Felder, G., & Fröhlich, J. (1992). Unified gauge theories in non-commutative geometry. *Physics Letters B*, 296(1), 109–116.

[Article Google Scholar](#)

- Chamseddine, A. H., Felder, G., & Fröhlich, J. (1993). Grand unification in non-commutative geometry. *Nuclear Physics B*, 395(3), 672–698.

[Article Google Scholar](#)

- Cheng, T.-P., Li, L.-F., and Cheng, T.-P.. (1974) Gauge theory of elementary particle physics. Oxford: Clarendon press, 1984. & Labo, Iyman. "Unity of all elementary-particle forces." *Physical Review Letters*.
- Christopher, L. H., & Elser, V. (1986). Quasicrystal structure of (Al, Zn) 49Mg32. *Philosophical Magazine B*, 53(3), L59–L66.

[Article Google Scholar](#)

- Churchland, P.S.. (2002) Brain-wise: Studies in neurophilosophy. MIT press
- Coldea, R., et al. (2010). Quantum criticality in an Ising chain: Experimental evidence for emergent E8 symmetry. *Science*, 327(5962), 177–180.

[Article PubMed Google Scholar](#)

- Connes, A.. (1988) Non-commutative geometry. Springer US.
- Conway, J. (1970). The game of life. *Scientific American*, 223(4), 4.

[Google Scholar](#)

- Conway, J. H., & Kochen, S. (2009). The strong free will theorem. *Notices of the AMS*, 56(2), 226–232.

[Google Scholar](#)

- El Naschie, M. S. (1998). Penrose universe and Cantorian spacetime as a model for noncommutative quantum geometry. *Chaos, Solitons & Fractals*, 9(6), 931–933.

[Article Google Scholar](#)

- Freeman, W. J., & Vitiello, G. (2006). Nonlinear brain dynamics as macroscopic manifestation of underlying many-body field dynamics. *Physics of Life Reviews*, 3(2), 93–118.

[Article Google Scholar](#)

- Fuller, R. B. (1975). *Synergetics*. New York: Macmillan.

[Google Scholar](#)

- Gautam, A., & Kak, S. (2013). Symbols, meaning, and origins of mind. *Biosemiotics*, 6(3), 301–310.

[Article Google Scholar](#)

- Grujić, P. V. (2011). The concept of fractal cosmos: III. Present state. *Serbian Astron Journal*, 182, 1–16.

[Article Google Scholar](#)

- Heisenberg, W.. (1971). "Physics and beyond: Encounters and conversations."
- Heyrovska, R. (2005). The Golden ratio, ionic and atomic radii and bond lengths. *Molecular Physics*, 103(6–8), 877–882.

[Article Google Scholar](#)

- Hooft, G.. (2000) "The holographic principle." arXiv preprint hep-th/0003004.
- Iovane, G., Laserra, E., & Tortoriello, F. S. (2004). Stochastic self-similar and fractal universe. *Chaos, Solitons & Fractals*, 20.3, 415–426.

[Article Google Scholar](#)

- Jaric, M.V., ed. (2012) Introduction to the Mathematics of Quasicrystals. Elsevier. p53.
- Kak, S. (1995). On quantum neural computing. *Information Sciences*, 83(3), 143–160.

[Article Google Scholar](#)

- Kaliteevski, M.A., et al. (2000) "Two-dimensional Penrose-tiled photonic quasicrystals: diffraction of light and fractal density of modes." *Journal of Modern Optics* 47.11: 1771–1778. & Amann, Anton, Lenz Cederbaum, and Werner Gans, eds. *Fractals, quasicrystals, chaos, knots and algebraic quantum mechanics*. Kluwer Academic Publishers, 1988, ch1?.
- Kellendonk, J. (1995). Noncommutative geometry of tilings and gap labelling. *Reviews in Mathematical Physics*, 7(07), 1133–1180.

[Article Google Scholar](#)

- King, C. C. (1996). Fractal Neurodynamics and quantum Chaos. *Fractals Brain Fractals Mind Adv Conscious Res*, 7, 179–233.

[Article Google Scholar](#)

- Kovacs, J., et al. (2012). "The Sum of Squares Law." arXiv preprint arXiv:1210.1446
- Kroto, H. W., et al. (1991). Crystal structure and bonding of ordered C 60. *Nature*, 353, 147–149.

[Article Google Scholar](#)

- Leibniz, M., (2004) 17, as quoted by Istvan Aranyosi. "[Chalmers's zombie arguments](#)" (draft ed.). Central European University Personal Pages.
- Lisi, A. G. (2007). An exceptionally simple theory of everything. *arXiv preprint, arXiv*, 0711.0770.

[Google Scholar](#)

- Livio, M.. (1991) The golden ratio: the story of Phi, the extra ordinary number of nature, art and beauty. Review, 2003. & Kroto, Harold W., et al. "Crystal structure and bonding of ordered C 60." *Nature* 353: 147–9. See also <http://io9.com/5985588/15-uncanny-examples-of-the-golden-ratio-in-nature>
- Maldacena, J. (2003). Eternal black holes in anti-de sitter. *Journal of High Energy Physics*, 04(2003), 021.

[Article Google Scholar](#)

- Penrose, R. (1979). "Set of tiles for covering a surface." U.S. Patent No. 4,133,152. 9 Jan.
- Penrose, R., & Hameroff, S. (2011). Consciousness in the universe: Neuroscience, quantum space-time geometry and Orch OR theory. *Journal of Cosmology*, 14, 1–17.

[Google Scholar](#)

- Perelman, C. C., Fang, F., & Irwin, K. (2013). Law of sums of the squares of areas, volumes and hyper-volumes of regular Polytopes from Clifford algebras. *AACA: Advances in Applied Clifford Algebras*, 23(4), 815–824.

[Article Google Scholar](#)

- Rosenblum B, and Kuttner F. Quantum enigma: Physics encounters consciousness. Oxford University Press, 2011.
- Rovelli, C. (2003). Loop quantum gravity. *Physics World*, 2003, 37–42.

[Article Google Scholar](#)

- Sadoc, J. F., & Mosseri, R. (1993). The E8 lattice and quasicrystals: Geometry, number theory and quasicrystals. *Journal of Physics A: Mathematical and General*, 26(8), 1789.

[Article Google Scholar](#)

- Smith, F. D. T.. (2004) E6, Strings, Branes, and the Standard Model. No. EXT-2004-031. and for his most recent model see <http://www.tony5m17h.net/1TSphysics.html>
- Snyder, H. (1947). http://prola.aps.org/abstract/PR/v71/i1/p38_1 "quantized space-time. *Physical Review D*, 67, 38–41.

[Article Google Scholar](#)

- Stanford Encyclopedia of Philosophy on Panpsychism (n.d.)
- Stapp, H.P. (2007) *Mindful universe: Quantum mechanics and the participating observer*. Springer.
- Susskind, L. (1995). The world as a hologram. *Journal of Mathematical Physics*, 36(11), 6377–6396.

[Article Google Scholar](#)

- Talbot, M. (1991). *The holographic universe*. London: HarperCollins Publishers.

[Google Scholar](#)

- Tegmark, M.. (2008) "Is “the theory of everything” merely the ultimate ensemble theory?." *Annals of Physics* 270.1 (1998): 1–51. Tegmark, Max. "The mathematical universe." *Foundations of Physics* 38.2: 101–150.
- The Elements of Physiology and Hygiene: A Text-book for Educational Institutions, (1868)by T.H. Huxley & W.J. Youmans. Appleton & Co. p. 178
- Ventegodt, S., et al. (2006). Human development III: Bridging brain-mind and body-mind. Introduction to “deep”(fractal, poly-ray) cosmology. *Scientific World Journal*, 6, 767–776.

[Article PubMed Google Scholar](#)

- Vitiello, G.. (2001) *My double Unveiled: the dissipative quantum model of brain*. John Benjamins Publishing, Vol. 32.
- Wang, H. (1961). Proving theorems by pattern recognition—II. *Bell System Technical Journal*, 40(1), 1–41.

[Article Google Scholar](#)

- Wheeler, J. A. (1986). Hermann Weyl and the unity of knowledge. *American Scientist*, 74, 366–375.

[Google Scholar](#)

- Wheeler, J.A.. (1990) "Information, physics, quantum: The search for links. In (W. Zurek, ed.) *Complexity, Entropy, and the Physics of Information*. Redwood City." CA: Addison-Wesley. Cited in DJ Chalmers,(1995) *Facing up to the Hard Problem of Consciousness*, *Journal of Consciousness Studies* 2.3: 200–19.
- Wilczek, F.. (2000), “What is quantum theory?”, *Physics today*, June.

- Winkelmann, S., et al. (2007) "An optimal radial profile order based on the Golden Ratio for time-resolved MRI." *Medical Imaging, IEEE Transactions on* 26.1: 68–76.
- Wolfram, S. (2002). *A new kind of science* (Vol. 5). Wolfram media: Champaign.

[Google Scholar](#)

- Xu, L., & Zhong, T. (2011). Golden ratio in quantum mechanics. *Nonlinear Science Letters B*, 1, 10–11.

[Google Schola](#)