A Timeless and Spaceless Quantum Theory of Consciousness

Jianfeng Li

ABSTRACT

In this article, a timeless and spaceless theory of consciousness in terms of quantum entanglement is proposed. It is shown that time and consciousness can be spontaneously emerged and, thus, be defined by separating a subsystem from a much bigger quantum system. The mass and relative distances are recognized as functions of the entanglement entropies between separated systems, and the classical background space can be restored in terms of these relative distances. We carefully examine the world with two coexisting consciousnesses. By equivalence principle of consciousness, the criterion of whether an entity is conscious or not is formulated implying that nowadays robots cannot be conscious. In addition, the relativity of intrinsic and extrinsic attributes of consciousness is discussed which might reveal the intricate connections between material and spiritual worlds at the deepest level. This relativity together with the new interpretation of mass further implies that consciousness has a negative mass, but this counterintuitive prediction needs to be verified by future experiments.

Key Words: quantum entanglement, relative distance, timeless and spaceless, Leibnizian philosophy

NeuroQuantology 2013; 3: 431-442

1. Introduction

The problem of consciousness, especially the Hard Problem of consciousness, might be the most challenging problem for human intelligence because it seems to resist all kinds of scientific investigations (Chalmers, 1995). In the past thirty years, various models and theories have been proposed to solve the problems regarding the consciousness (de Sousa, 2013). But Chalmers (1995) points out that these theories including quantum theory (Smith, 2006;2009) can at best solve the easy problems of consciousness but always fall short of the target when facing the hard problem. Therein, he defines the easy problems as those seem directly susceptible to the standard methods of cognitive science, while the hard problems are those that seem to resist those methods. In particular, the hard problem is usually referred to the problem of conscious experience. To solve the hard problem, he proposes a double-aspect theory of information (or dual aspect science (Hales, 2009)) which postulates that physical entities have both extrinsic and intrinsic properties and there are some mapping relationships between these two different kinds of attributes. Note that intrinsic and extrinsic properties roughly correspond to implicate and explicate orders in Bohm's theory of mind and matter (1990).

However, the intrinsic and extrinsic properties of a physical entity should be relative depending on who is the observer. For example, in my eyes my conscious experience and some properties of the external world (e.g. shape of a table) are extrinsic or explicate; but for other people, my conscious experience will become intrinsic because they will never know what I am thinking about at the moment. We

Corresponding author: Jianfeng Li
Address: The State Key Laboratory of Molecular Engineering of Polymers, Department of Macromolecular Science, Fudan University, Shanghai 200433, China
✉ lijf@fudan.edu.cn
Relevant conflicts of interest/financial disclosures: Author thanks NSFC (Grant No. 21104010) for financial support.
Received July 14, 2013. **Accepted** August 17, 2013.
*e*ISSN 1303-5150

believe that the relativity between intrinsic and extrinsic properties holds the key to eventually solving the Hard Problem as well as to bridging the gap between dualism and monism.

This article aims to develop a timeless and spaceless quantum theory of consciousness that can account for this relativity of intrinsic and extrinsic properties of consciousness. We will specially focus on how to define consciousness, time, relative distance and mass in this prespacetime theory.

2. Four Principles and Outline of the theory

Our theory of consciousness is based on the following four principles (necessary explanations are also given):

(i) *The quantum superposition principle is assumed to be universally correct.* This is the only principle of quantum theory that can be employed in the current theoretical construction for a prespacetime theory. There are some other quantum theories of consciousness (Smith, 2006; 2009; Hameroff, 1995; 2007; Hameroff and Penrose, 2003; Mensky, 2013), but seldom is timeless and spaceless. Further, it should be noted that in many respects our theory can be seen as an extended theory of Mensky's extended-Everett-interpretation (EEI) theory (2013; 2011; 2007) but our theory is without time and space. We expect this retreat can present us a broader view of consciousness.

(ii) *Any physical property, that is accessible by physical means, is extrinsic, and it is nothing more and nothing less than the relations of the entity to the rest of the world.* This principle is actually based on Leibnizian philosophy regarding perception about monads (Barbour, 2003). But our principle does not deny the existence of the intrinsic properties or conscious experience, which cannot be detected by any physical measurements. This principle is obviously useful in restoring the concepts of time, distance and mass in a timeless and spaceless theory of consciousness: they are indeed merely the relations specified by entanglement entropies within the framework of quantum theory. Entanglement entropies are the only quantities that have physical meanings for a timeless and spaceless quantum theory.

(iii) *Anthropic principle.* We postulate the fact that the consciousness can feel or memorize the past largely determines the configuration of the emergent time. The requirement that the neighbouring instant worlds should be closed to each other is indeed a premise for a consciousness to memorize its past, which is the key to interpret anthropic principle in the current context. In some sense, Barbour's time capsule and best-matching method (1999) can be seen as another way of implementation of the anthropic principle .

(iv) *Equivalence Principle of Consciousness.* Just like the equivalence principle in general relativity, this principle is derived from Leibniz's principle of sufficient reason (Barbour, 2003) and it can be understood from the following two levels. First, any description of the universe dooms to be the description of the universe in the eyes of some consciousness. Second, the physical rules of the universes in the eyes of two different normal consciousnesses should be basically the same, though specific contents may differ and in particular the extrinsic and intrinsic properties may exchange for two different consciousnesses' worlds.

Based on the above principles, a super quantum state will be firstly introduced in this work. By separating a subsystem from this quantum state, we found that it is possible to define time through properly ordering the orthonormal states of the subsystem in a sequence. The Hamiltonian of the sequence can be also defined during this process, but is found to be not unique for each sequence. The requirement of the quasi-uniqueness of Hamiltonian (or physical rules) leads to the quasi-uniqueness of the universe that accompanied by the subsystem. The subsystem can be then recognized as the premature consciousness. By requiring the premature consciousness to be able to remember the past, it is found that the time has to branch into the future and, therefore, free will is emerged. The relative distance of systems will be defined in terms of entanglement entropies between these systems. It will be easily recognized that the mass and gravity have been implicitly contained in the definition of relative distance. The usual background space can be restored from the relative-distance configuration (Barbour, 1974). We discuss two consciousnesses coexisting in the same universe. By exchanging the viewpoints of these two consciousnesses together with equivalence principle of consciousness, we are able to determine whether a system is conscious or not. In particular, it is found that the nowadays computer cannot be conscious.

Finally, the relativity of intrinsic and extrinsic properties of consciousness will be extensively discussed within the above theoretical framework. In particular, we introduce the concept of *induced extrinsic* which might play an important role in the understanding of phenomena related to consciousness. Based on this relativity and the new interpretation of mass, a very bold prediction is presented claiming that consciousness might have negative mass: the entity bearing consciousness will lose some weight during the formation of consciousness. Unluckily, this prediction is found to be difficult to verify in practice.

3. A Quantum State Contains Everything But Tells Nothing

A super quantum state is firstly introduced

$$|D\rangle = \sum_d D_d |d\rangle \qquad (1)$$

where $|d\rangle$ are N_d orthonormal states and D_d the complex numbers. It will be convenient to name this super quantum state *Dao*.

At the moment $|D\rangle$ is just an ordinary quantum state in a N_d dimensional Hilbert space. It is timeless and spaceless or in other words time, space, observables, Hamiltonian and mass is not pre-defined within this quantum state. *Dao contains everything but tells nothing*.

Note that this super quantum state Dao is similar to Bohm's *wholeness* (1980), Mensky's *pure existence* (2013) and the wave function in Wheeler-DeWitt's equation (DeWitt, 1973). However, there are differences between them: time exists in both Bohm's wholeness and Mensky's pure existence while time is absent in the wave function in Wheeler-DeWitt's equation but it is equipped with space.

4. Permutation of Orthonormal States and Emergence of Time

4.1 Separation of a system from a super quantum state

In Mensky's EEI theory (2013; 2011; 2007), separation generates consciousness. It plays a similar role in our theory. However separation here not just produces consciousness but also gives rise to all physical entities.

If the dimension of the super quantum state Dao is a composite number, e.g. $N_d = N_m \times N_w$, then we can decompose it as an entangled state of a composite system of M with dimension N_m and W with dimension N_w (or product of two pure states). Generally, such composite system can be expressed as

$$|D\rangle = \sum_k \sum_j C_{kj} |M_k\rangle |W_j\rangle. \qquad (2)$$

The coefficients C_{kj} should be equal to D_d. But there are $N_d!$ different ways to assign (map from) $\{D_d\}$ to $\{C_{kj}\}$ if D_d are different from each other. Since the order of the basis of M is not important, there are actually $N_d!/N_m!/N_v!$ different decompositions. If C_{kj} can be expressed as $C_k C_j$, then M and W are two separable, pure quantum states.

A case study. Consider a six dimensional super quantum state

$$|D\rangle = \frac{3}{2}|D_1\rangle + |D_2\rangle + \frac{1}{2}|D_3\rangle + 3|D_4\rangle + 2|D_5\rangle + |D_6\rangle$$

(not normalized). There are $6!/3!/2! = 60$ ways to separate a 2D subsystem from it. Two typical ways are

$$|D\rangle = \frac{3}{2}|M_1\rangle|W_1\rangle + |M_1\rangle|W_2\rangle + \frac{1}{2}|M_1\rangle|W_3\rangle$$
$$+3|M_2\rangle|W_1\rangle + 2|M_2\rangle|W_2\rangle + |M_2\rangle|W_3\rangle$$

and

$$|D\rangle = 3|M_1'\rangle|W_1'\rangle + |M_1'\rangle|W_2'\rangle + \frac{1}{2}|M_1'\rangle|W_3'\rangle$$
$$+\frac{3}{2}|M_2'\rangle|W_1'\rangle + |M_2'\rangle|W_2'\rangle + 2|M_2'\rangle|W_3'\rangle .$$

It is clear that the first one is indeed the product of two pure states, i.e., $(|M_1\rangle + 2|M_2\rangle)(\frac{3}{2}|W_1\rangle + |W_2\rangle + \frac{1}{2}|W_1\rangle)$; while the second one is a mixed state with entanglement entropy $S \approx 0.36$.

In Barbour's relative-distance machian theory (1974) the universe becomes nontrivial unless it has more than three particles. In our theory, the universe becomes nontrivial unless $|D\rangle$ has more than four dimensions. If $|D\rangle$ is two or three dimensional then there is only one way to decompose $|D\rangle$ as 1x2 (or 1x3) which further corresponds to two particles with one dimensional and two dimensional freedoms, respectively. The point is that these two particles must be two pure states, which means they are totally unaware of each other, so it is actually two trivial universes each with only one particle. If the universe is four dimensional, then the two particles each with two dimensional freedom will be quantum entangled with each other. If they are two classic particles, the universe is still trivial according to Barbour's argument (1974). But there are three different ways to divide the four dimensional quantum state into two particles. Though for each way there are only two particles, the pairs of particles in different ways are correlated with each other to form a nontrivial universe (even though it is still a very boring universe).

We emphasize that *separation or decomposition of a bigger quantum state into smaller ones is a very very important mathematical skill in our theory* which turns *nothing* into *everything*.

4.2 Emergence of time

For a Dao equipped with a decomposition, we shall name W the *universe* while M will be recognized as *consciousness* if it satisfies certain conditions. The first condition is $N_w \gg N_m \gg 1$, and M and W are highly entangled. If M and W are unentangled, then M can simply not be able to feel W. N_m will be used to build the time so it cannot be small; otherwise M will have a short, miserable cyclic life. Other conditions will be given one by one in the following context.

Consider a sequence $P_M = \{|M_1\rangle, |M_2\rangle, \ldots |M_n\rangle\}$ which is a subset of M with $n \le N_m$ and define a sub-Dao based on this permutation

$$|D_{P_M}\rangle = \sum_k |M_k\rangle \sum_i C_{kj} |W_j\rangle, \tag{3}$$

which satisfies $\langle M_k | [|D\rangle - |D_{P_M}\rangle] = 0$. Note that unless $n = N_m$, the sub-Dao $|D_{P_M}\rangle$ will not equal to Dao $|D\rangle$. One can always normalize $\sum_j C_{kj} |W_j\rangle$ for each k and put the normalization factor out of the summation. In this way we rewrite the above expression as

$$|D_{P_M}\rangle = \sum_k d_k |M_k\rangle \sum_j C_{kj} |W\rangle. \tag{4}$$

where the normalization factor d_k is real and $\sum_j C_{kj} C^*_{kj} = 1$ for each k. It is convenient to define $|\bar{W}_k\rangle = \sum_j C_{kj} |W_j\rangle$ as an *instant NOW* (Barbour, 1999).

We have to further assume that the world is changing smoothly, i.e. $|||\bar{W}_{k+1}\rangle - |\bar{W}_k\rangle||^2 = \sum_j ||C_{(k+1)j} - C_{kj}||^2 \ll 1$ for all k. Otherwise it will soon be clear that time and consciousness cannot be defined. In the discussion section, we show that (not very rigorous) for $N_w \gg N_m \gg n$, this condition can be easily satisfied. Actually, the role of this condition is similar to that of the best matching in Barbour's timeless-world theory (1999).

Let's check whether we can find a $N_w \times N_w$ matrix $H_{jj'}$ that satisfies

$$C_{(k+1)j} = \sum_{j'} (\delta_{jj'} + iH_{jj'}\varepsilon) C_{kj} \tag{5}$$

for all k and j, where ε is a small constant real number and $\delta_{jj'} = 1$ if $j = j'$, $\delta_{jj'} = 0$ otherwise. First the constraints of $\sum_j C_{kj} C^*_{kj} = 1$ and $\sum_j C_{(k+1)j} C^*_{(k+1)j} = 1$ require H to be a hermitian. Second, there are N_w^2 unknowns (the number of elements of the matrix) and $N_w \times n$ equations. Therefore, if $n \rangle N_w$, i.e. the order of permutation is bigger than the freedom of the universe, there will be more equations than the unknowns and we might not be able to find a matrix H. However, by the assumption, the freedom of the universe should be much bigger than that of M, i.e. $N_w \gg N_m$. Therefore $N_w \gg n$ which indicates we have more than one solutions. We denote the set of the solutions for the permutation P_M as $\{H\}_{P_M}$. For each solution $H_l \in \{H\}_{P_M}$, the evolution of the universe can be expressed as

$$\sum_j C_{kj} |W_j\rangle = e^{iH_l t} \sum_{j'} C_{0j'} |W_{j'}\rangle \tag{6}$$

where

$$t \equiv k\varepsilon \tag{7}$$

should be recognized as the time and H_l as one of Hamiltonians of P_M.

Note that the situation here is similar to that of the quantum gravity (QG) theory (Kiefer, 2007) and Barbour's timeless-world theory (1999). In Barbour's theory (1999), the time is defined through the procedure of best matching of neighbouring instant NOWs. In our theory, time emerges from the separation of Dao into two subsystems: consciousness and universe; and the space and mass can be defined in the further separation of universe into numerous subsystems.

5. Quasi-Uniqueness of Hamiltonian and Emergence of Premature Consciousness

In the preceding section, we may have many choices of Hamiltonian (or physical rules) for a single universe which apparently contradicts the intuition about our universe. Therefore, we require that $H_i \approx H_j$ for any two Hamiltonian belong to $\{H\}_{P_M}$ and this is the condition of Quasi-uniqueness of Hamiltonian. This can be achieved in the following way. If for each k, only about n coefficients C_{kj} are not close to zero, the dimension of the matrix H can be effectively thought as $n \times n$. Other elements of H are not really vanished, but they are not very important in determining the evolution path of the universe W.

If the Hamiltonian of the universe satisfies this quasi-uniqueness condition, we call the permutation P_M as a *premature consciousness*. In other word, if P_M is a premature consciousness, the number of non-zero elements of C_{kj} must be about the time segments of the consciousness, n.

For each k or $t = k\varepsilon$, a state $|W_j^*(k)\rangle$ or $|W(t)\rangle$ with maximum $||C_{kj}||$ for all j is selected to represent the universe state at time t. In this way, we can plot $||C_{kj}||$ roughly along the $|W(t)\rangle$ axis. Note that the plot of $||C_{kj}|| \sim |W(t)\rangle$ can be roughly seen as the wave function of the instant NOW k spanning on the basis $\{|W(t)\rangle\}$. Because $\sum_j ||C_{(k+1)j} - C_{kj}||^2 \ll 1$, therefore the profiles of C_{kj} and $C_{(k+1)j}$ must be close to each other. This actually requires that the profile of $||C_{kj}||$ for an instant NOW should take the shape as shown in Figure 1 (if the universe is symmetric about the time).

6. Implications of Branching of Time and Human Consciousness

Usually a sequence $P_M = \{|M_1\rangle, |M_2\rangle, \ldots |M_n\rangle\}$ with length n will not exhaust all the orthonormal states of the subsystem M.

Actually, the orthonormal states of M can be possibly divided into several unconnected chains as shown in Figure 2, where each circle represents one orthonormal state. Every two neighbouring states satisfy the condition of $\sum_j ||C_{(k+1)j} - C_{kj}||^2 \ll 1$.

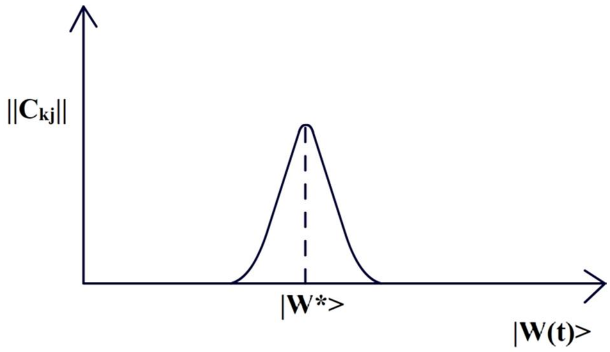

Figure 1. The schematic illustration of the profile of $||C_{kj}|| \sim |W(t)\rangle$ at time $t = k\varepsilon$.

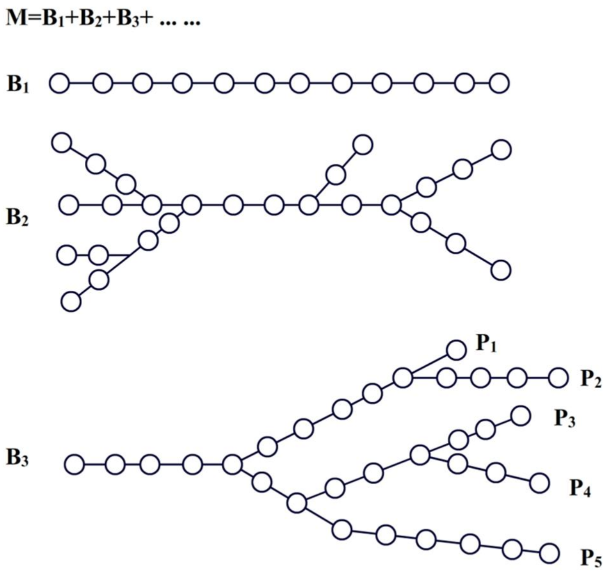

Figure 2. The schematic illustration of the branching of *M*.

Actually, only the future-branching case (B_3 in Figure 2) reflects the reality. In this case, the shape of the $||C_{kj}||$ is half-peak where the future half peak almost vanishes (see Figure 3). The reason for this is that human consciousness is able to aware of the past other than future. Intuitively, the different future branches cancel out the future half peak because C_{kj} can be positive or negative. Unfortunately, we are not able to provide a rigorous proof here and the mechanism of awareness is not clear at the moment. But one thing is sure, for the consciousness to be aware of (or memorize) the past, the information of the past must be stored in the coefficients C_{kj} in one way or another. On the other hand, if the information of the future is also manifestly contained in C_{kj}, then the consciousness may find some way to 'remember' the future.

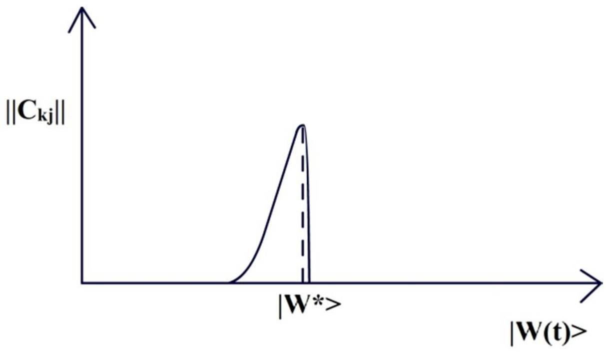

Figure 3. The schematic illustration of the half-peak shape of $||C_{kj}|| \sim |W(t)\rangle$ at time $t = k\varepsilon$.

Therefore, we have defined the human consciousness precisely in terms of quantum states: For a quantum state (Dao) that can be expressed as

$$|D\rangle = \sum_m d_m |M_m\rangle \sum_j C_{mj} |W_j\rangle$$

where $\{|M\rangle\}$ are N_m orthonormal states, if there exists a sequence $P_M = \{|M_1\rangle, |M_2\rangle, \ldots |M_n\rangle\}$ with $M_k \in \{M\}$ that satisfies

(i) $\sum_j ||C_{(k+1)j} - C_{kj}||^2 \ll 1$

with $|D_{P_M}\rangle = \sum_k d_k |M_k\rangle \sum_w C_{kw} |w\rangle$

and $\langle M_k|(|D\rangle - |D_{P_M}\rangle) = 0$ for all k;

(ii) $N_w \gg n$;

(iii) The profile of the curve $||C_{kj}(t)|| \sim |W(t)\rangle$ is half-peak. For each k, except for n none-zero C_{kj}, all other $||C_{kj}||$ are extremely close to zero.

Then we define the sequence P_M as the human consciousness and $t = k\varepsilon$ as the time with ε a constant that should be related with the Planck constant h. In the following context, we may just set $\varepsilon = 1$ for simplicity.

If human consciousness is defined in this way, then it can branch into future which further indicates that human consciousness has a free will. If there is only one future branch ahead, then the free will would be a fake one (or an elusion), because everything in the future has been exactly scheduled and no real choice we can make for the future. In this case the profile of $||C_{kj}(t)|| \sim |W(t)\rangle$ will be symmetric about k (Figure 1), which means we will also 'remember' the future! But anything we remember should be defined as the 'past'. Therefore, in this single-branch world the consciousness might not be able to sense the time passing. Because for the consciousness at $t = k$, both $|M_{k-1}\rangle$ and $|M_{k+1}\rangle$ are its past. More probably, the consciousness in this special world cannot experience anything.

If the world is branching into both future and past, then the profile of $||C_{kj}(t)|| \sim |W(t)\rangle$ is a very narrow peak indicating that M cannot remember anything and thus cannot experience anything too. According to the common sense of the consciousness, these two types of premature consciousness are not conscious.

From the above discussion, we learn that the time symmetry is broken because *human consciousness can only 'remember' the past and this is what we define the past; and time can only lapse into the future that the consciousness cannot 'remember' and this is what we define the future*. Note that the anthropic principle has been employed in the above discussions.

There is still one problem about the future branching universe. If there is only one physical (dynamical) rule, how can the past evolves into future in different branches? It is simply because there are more than one physical rules (or Hamiltonian). For different branches, $H_i \neq H_j$ although they share the same past. But we may have $H_i \approx H_j$. Actually even for the same branch, there are also more than one Hamiltonian that are different from each other but closed to each other according to the quasi-uniqueness properties of Hamiltonian.

Indeed in the above discussions, we have resolved the inconsistency between the unitary evolution and wave function collapse in quantum mechanics. The key is that the unitary evolution or the Hamiltonian is not pre-equipped and is not exactly unique. We admit that if the evolutionary rule or Hamiltonian is unique then there is probably no wave function collapse at all. Therefore, our interpretation of quantum mechanics is different from Everett's many-world interpretation (Everett, 1957) and the decoherence interpretation (Joos, 2003; Kiefer, 2007) in this respect.

7. Two Consciousnesses

Consider a consciousness $P_M=\{|M_1\rangle,|M_2\rangle...|M_k\rangle\}$. It is assumed that there is another consciousness $\tilde{M}$ residing in the accompanied universe of M, i.e., $|W\rangle=|\tilde{M}\rangle|W'\rangle$. The sub-Dao $|D_{P_M}\rangle$ is rewritten as

$$|D_{P_M}\rangle=\sum_k d_k|M_k\rangle[\sum_{k',j}C_{kk'j}|\tilde{M}_{k'}\rangle|W'_j\rangle]. \quad (8)$$

If the consciousness $\tilde{M}$ is deeply entangled with W', then the quantum state $\sum_{k',j}C_{kk'j}|\tilde{M}_{k'}\rangle|W'_j\rangle$ will also become a Dao for each k which indicates that $\tilde{M}$ might form a consciousness for a single k. What this really means is that when you are staring at your conscious friend, your friend's consciousness might have been through a whole life in a very narrow time span ($\Delta t=\varepsilon$). To avoid this possibly odd situation, it is required that $\tilde{M}$ is only slightly entangled with W'. Therefore, $|D_{P_M}\rangle$ can be approximately written as

$$|D_{P_M}\rangle=\sum_k d_k|M_k\rangle\sum_{k'}C_{kk'}|\tilde{M}_{k'}\rangle\sum_j C_{kj}|W'_j\rangle. \quad (9)$$

Note that the states $\sum_{k'}C_{kk'}|\tilde{M}_{k'}\rangle$ cannot be totally orthogonal with each other. Otherwise, there exists one-to-one mapping from $|M_k\rangle$ to $\sum_{k'}C_{kk'}|\tilde{M}_{k'}\rangle$ and therefore M_k and $\sum_{k'}C_{kk'}|\tilde{M}_{k'}\rangle$ are actually the same consciousness. This also indicates that M cannot be totally entangled with $\tilde{M}$, otherwise they will be the identical consciousness.

Let's focus on the shape of $||C_{kk'}||$. Suppose it is also half-peak on the past side just as that of $||C_{kj}||$. The sub-Dao $|D_{P_M}\rangle$ can be also written as, in the eyes of the consciousness $\tilde{M}$

$$|D_{P_M}\rangle=\sum_{k'}d'_{k'}|\tilde{M}_{k'}\rangle\sum_k C_{k'k}|M_k\rangle\sum_j C_{kj}|W'_j\rangle \quad (10)$$

where $C_{k'k}=C_{kk'}$ (if we assume $d'_{k'}\approx d_k$). Then it is not difficult to show that $||C_{k'k}||$ is half-peak on the future side which is obviously different from that of $||C_{kk'}||$ (Figure 4). Therefore, the equivalence principle of consciousness is violated and the shape of $||C_{kk'}||$ cannot be half-peak. So it should be a full peak but narrow enough otherwise the consciousness is able to know the future of the other consciousness. We denote the width of this peak as σ and obviously $1\ll\sigma\ll n$.

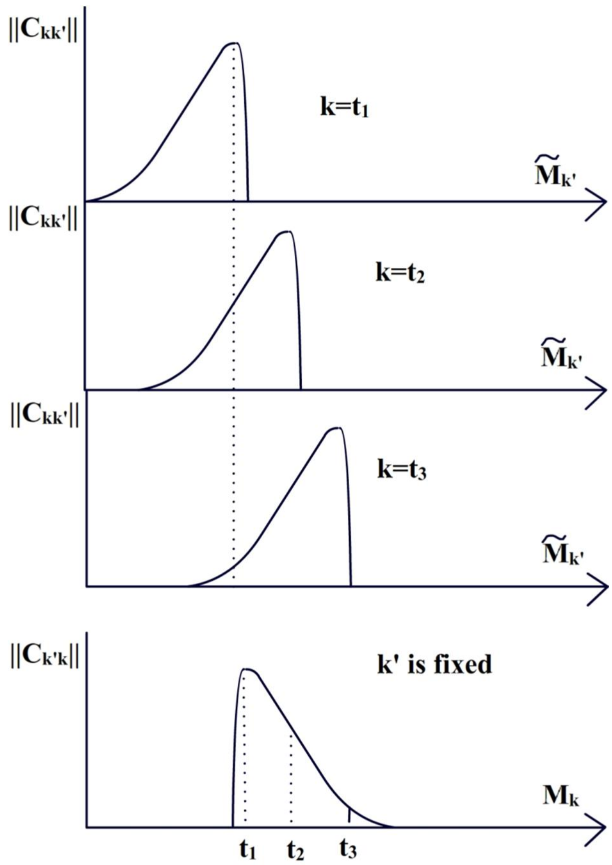

Figure 4. The schematic illustration of the difference between $||C_{kk'}||\sim\tilde{M}_{k'}$ and $||C_{k'k}||\sim M_k$.

We emphasize that exchanging the viewpoints of two consciousness in Dao is another important mathematical technique in the theory of consciousness.

8. Is Robot Conscious

Consider a robot that is running by some complicated program. We abstract the robot using a sequence of states $\{|R_1\rangle,|R_2\rangle,...|R_r\rangle\}$ that are orthogonal with each other. Our primary conclusion is that a nowadays robot cannot be conscious.

(1) For the robot to be conscious, the integer r has to be extraordinary big in the first place because the robot has to use these states to reconstruct the seemly continuously flowing

time. But for nowadays computer or robot, r cannot be comparable with n or N_m.

(2) More importantly, if robot is conscious, then $|D_{P_M}\rangle$ should be written as

$$|D_{P_M}\rangle = \sum_k d_k |M_k\rangle \sum_l C_{kl} |R_l\rangle \sum_j C_{kj} |W'_j\rangle. \quad (11)$$

According to the discussion in the above section, for each time being $t=k$, the robot state is actually a superposition state of roughly σ states, which is certainly against the deterministic nature of the programming of the robot.

(3) On the other hand, there might be several k (several instant NOWs) that correspond to the same robot state. Then in the view of robot, the world and the consciousness M are both highly superpositional. The robot might have a very difficult time to discriminate or experience these blur worlds in order to develop any meaningful concepts. For example, suppose $k=1,2,3$ corresponds to a given l, then for R_l, we have

$$|R_l\rangle[d_1|M_1\rangle + d_2|M_2\rangle + d_3|M_3\rangle][|W(1)\rangle + |W(2)\rangle + |W(3)\rangle].$$

(4) Finally, the deterministic nature of the robot indicates that it has no free will. As discussed in the section V, no free will actually means no sensation and no consciousness.

Nevertheless, our theory does not turn down the possibility that a robot can be clever or intelligent or even creative. In addition, it should be noted that robots based on other techniques can be conscious because human being is an example.

9. Distance, Gravitation and Mass of Consciousness

For an instant NOW, consciousness M and universe W are two pure states. As in Figure 5, the pure state W can be further decomposed into E_1, E_2 and W' expressed as $|W\rangle = \sum_{e_1e_2w} C_{e_1e_2w} |E_{e_1}\rangle |E_{e_2}\rangle |W_{w'}\rangle$. The entries of the density matrix ρ_{E_1} are calculated as $\rho_{e_1e'_1} = \sum_{e_2,w} C_{e_1e_2w} C^*_{e'_1e_2w}$, and those of ρ_{E_2} are obtained similarly. We define $\rho_{e_1e_2e'_1e'_2} = \sum_w C_{e_1e_2w} C^*_{e'_1e'_2w}$ as the entries of $\rho_{[E_1E_2]}$. Based on these density matrices, one can easily obtain the corresponding entanglement entropies S_{E_1}, S_{E_2} and $S_{[E_1E_2]}$ which actually reflect the relations among these entities. The entanglement entropy between entities E_1 and E_2 is defined as $S_{E_1E_2} \equiv (S_{E_1} + S_{E_2} - S_{[E_1E_2]})/2$.

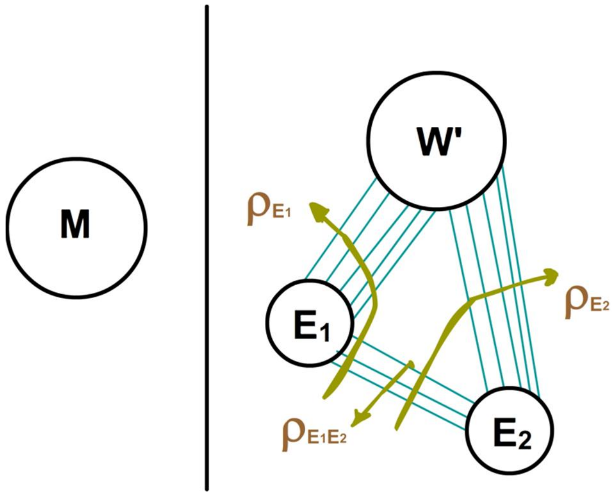

Figure 5. For an instant NOW, E_1, E_2 and W' consist of a pure quantum state. The distance between E_1 and E_2 is completely determined by the relations of E_1 and E_2 with respect to W' specified by the density matrices ρ_{E_1}, ρ_{E_2} and $\rho_{E_1E_2}$.

It is expected that any physical quantities, such as relative distance, related with E_1 and E_2 should be defined as functions of these three entropies (see principle (ii) in Section 2). It is easy to see that bigger $S_{E_1E_2}$ corresponds to smaller relative distance and therefore the simplest distance function that one can make up takes the following form

$$d_{E_1E_2} = \frac{S_{E_1} S_{E_2}}{S_{E_1E_2}}. \quad (12)$$

By exchanging the positions of $d_{E_1E_2}$ and $S_{E_1E_2}$, it is not difficult to recognize that the above equation is also the equation of gravitation where $S_{E_1E_2}$ is the gravity potential, S_{E_1} and S_{E_2} are actually the masses of entities E_1 and E_2. In addition, it has following implications:

(i) Mass actually measures the connections of an entity to the rest of the world. According to this definition, mass is approximately additive. For example, $m_{E_1E_2} = m_{E_1} + m_{E_2} - 2S_{E_1E_2} \approx m_{E_1} + m_{E_2}$ since normally $m_{E_i} \gg S_{E_1E_2}$.

(ii) Gravity appears inevitably as long as there is relative distance between systems.

(iii) The mass of rest of world with respect to E_1 can be computed as exactly the same as the mass of E_1. For example, the mass of a universe excluding an electron is actually the mass of that electron. In this connection, we can safely set the total mass of the universe to be zero. It should be noted that this result is coincident with that of Wheeler-DeWitt equation in quantum gravity theory which actually states the total energy of the universe is zero.

Based on the implication (i), we postulate that during the formation of a consciousness, the brain will lose some weight, i.e., *consciousness has a negative mass*! This counterintuitive hypothesis can be easily understood by examining the formation of consciousness. It is expected that during the formation of consciousness lots of elementary particles behave as a whole and some connections of the particle to the rest of the world are transformed into intrinsic properties (or inner freedoms) of the consciousness. The disappearing of the connections will cause a weight loss of the consciousness according to implication (i). Obviously, this hypothesis violates the conservation law of mass (energy). According to Noether's theorem, the conservation law of mass (energy) originates from the differentiable symmetry of action with respect to time and but there is no such symmetry in the current theory. Instead, the conservation law of mass (energy) is approximately correct: the density matrix ρ_{E_1} is actually tracing over (or averaging over) the rest of world, if the freedom of the rest of world is much bigger than that of E_1 than the entanglement entropy S_{E_1} will approximately remain unchanged.

It is possible to estimate the mass of consciousness. If for each instant NOW, consciousness M will split into two copies and total possible length of M is N_m then total freedom of M will be roughly $2^{N_m+1}-1$ (other branches have been ignored). The entropy will be $S \sim N_m \ln 2 = -\Delta mc^2 T/\hbar = -mc^2 nt_p/\hbar$ with c the light speed, nt_p the life span of the consciousness M and t_p the Planck time. Therefore,

$$|\Delta m| \sim \frac{N_m \hbar \ln 2}{nt_p c^2} \sim 1.51\times 10^{-5}\frac{N_m}{n} gram. \tag{13}$$

If N_m/n is big then this amount of mass loss can be detected by an extremely delicate experiment. Hopefully in the near future, we can really test this hypothesis. It will be exciting if the hypothesis is true because it is so counterintuitive. Further, it will provide a very simple physical means to determine whether a robot is conscious or not: just put it on the scale!

10. Relativity of Intrinsic and Extrinsic

For each instant *NOW*, M and W are disentangled and it is impossible for the consciousness M to feel about universe W. But if a time span $t=t_0 \to t_0+\Delta t$ is taken into consideration, all instant NOWs in this time span can be added together to form a bigger pure state $\sum_{k_1}^{k_2} d_k |M_k\rangle \sum_i C_{ki}|W_i\rangle$ with $t_0 = k_1 t_p$ and $t_0+\Delta t = k_2 t_p$ (Figure 6). For a given consciousness M and a given time span Δt_M, the extrinsic properties is strictly defined as those properties that can be described by the density matrices directly related to M. Herein, perception of M and inner structures are extrinsic while relations inside W (for example S_{E_1}) are intrinsic. If all relations inside W are intrinsic, then it is obviously impossible for M to perceive mass, distance and other properties of matters which are actually the relations in W. This is certainly not true in the real world. Actually, these properties can be deduced from the perception of M specified by the density matrix $\rho_{MW}(\Delta t_M)$ for different time span Δt_M. Accordingly, we introduce the concept of *induced extrinsic properties* as those properties inside W that will approximately remain unchanged when time span Δt_M approaches zero. Note that all physical quantities are actually induced extrinsic, which can be approached by scientific investigations. We should take *extrinsic* and *induced extrinsic* as simply *extrinsic*, and others as *intrinsic*. For example, for consciousness M the perceptions of consciousness $\tilde{M}$ are intrinsic (Figure 7).

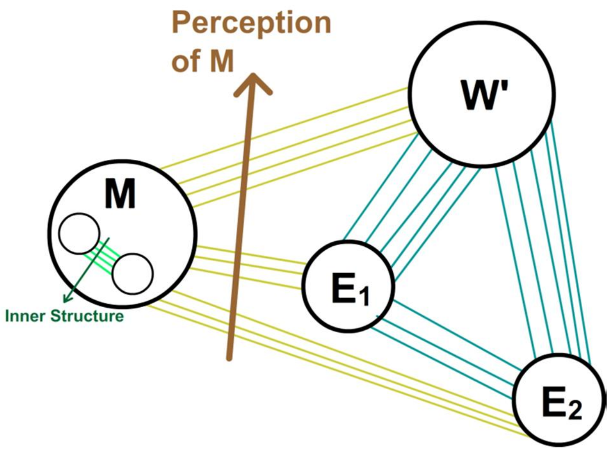

Figure 6. Several instant NOWs with time span $\Delta t_M > 0$ are added together to form a bigger pure state. For a given consciousness M and a given time span Δt_M, the extrinsic properties is strictly defined as those properties that can be described by the density matrices directly related to M. Herein, perception of M and inner structures are extrinsic.

More discussions about relativity of intrinsic and extrinsic properties of consciousness are presented as follow: (i) The time span Δt_M is usually about the order of 1 second. (ii) If induced extrinsic is taken into account then the border between intrinsic and extrinsic is actually a little blur, which may correspond to the uncertainty principle in quantum mechanics (note that we did not include this principle into our theory at the beginning). (iii) Because the induced extrinsic properties will roughly remain unchanged when time span approaches zero, so it is sufficient to study the induced extrinsic properties only by investigating some instant NOW where M and W are two disentangled pure states (see the preceding section for example). (iv) For a given t, we can define a quantum state $|D_{W^*}\rangle = |W^*(t)\rangle(\sum C_m |M_m\rangle)$ such that $\langle W^*(t)|(|D\rangle - |D_{W^*}\rangle) = 0$, where the definition of $|W^*\rangle$ is referred to Figure 3. The inner structures of consciousness M at time t are encoded in the quantum state $\sum_m C_m |M_m\rangle$. Note that this state is dual to the $\sum_i C_{ki} |W_i\rangle$. Therefore, if you are interested in material world you should focus on the quantum state $\sum_i C_{ki} |W_i\rangle$. On the other hand, if you are abstracted by the spiritual world then $\sum_m C_m |M_m\rangle$ will be a better choice.

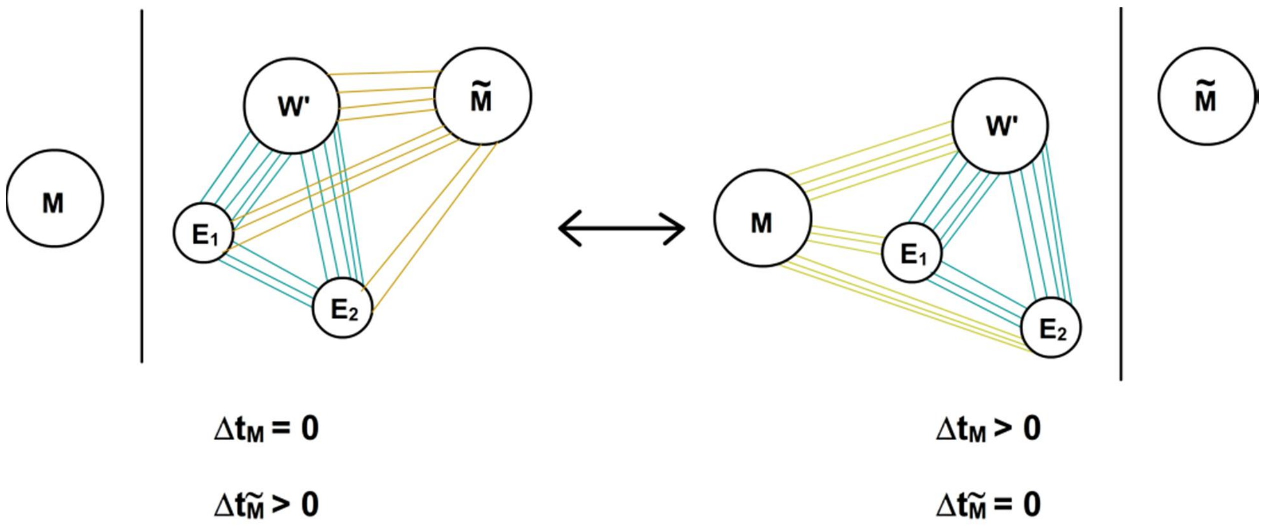

Figure 7. The induced extrinsic properties is defined as those that can approximately remain unchanged when Δt_M is approaching zero. Obviously, perception of $\tilde{M}$ is not induced extrinsic for M. This figure also expresses the equivalence principle of consciousness.

11. Discussion

11.1 Quasi-Continuous Condition

In section 4.2, we have imposed a quasi-continuous condition for neighboring W's, i.e. $\sum_j ||C_{(k+1)j} - C_{kj}||^2 \ll 1$. This condition actually imposes constraints on the coefficients D_d and the decomposition of Dao. But it will be nice if this condition can be automatically (or naturally) satisfied.

It will be natural to assume that the coefficients D_d of $|D\rangle$ are a series of random complex numbers. In following context, we will give a simple illustration (NOT a proof) about how to generate these quasi-continuous

C_{kj} from this setting. Firstly, we should note that $N_d = N_w \times N_m$ with N_d, N_w and N_m dimensions of Dao $|D\rangle$, universe $|W\rangle$ and consciousness M, respectively, and $N_d \gg N_w \gg N_m \gg 1$. We assume that N_w coefficients of $\{D_d\}$ have been already assigned to C_{0j}. The problem is reduced to find another N_w coefficients in the left $(N_m-1)N_W$ coefficients of $\{D_d\}$ to assign to C_{1j} in order that $\sum_j ||C_{1j}-C_{0j}||^2 \ll 1$. C_{0j} can be thought as a point on a $2N_W$ dimensional unit sphere with area $\sim\pi^{N_w}$. There are $((N_m-1)N_w)!/N_w!$ different ways to choose C_{1j} from $\{D_d\}$, which corresponds to $((N_m-1)N_w)!/N_w!$ different points on the $2N_W$ dimensional unit sphere. Because D_d are random complex numbers, so these points can be roughly thought as randomly distributed on the sphere and on average every point occupies an area of $\sim\pi^{N_w} N_w!/((N_m-1)N_w)!$ which is extremely small. Therefore, it is possible to find a point that is sufficiently close the given point C_{0j}. Let the point be C_{1j}. Then we have $\sum_j ||C_{1j}-C_{0j}||^2 \ll 1$. Repeating the above procedure, we can find C_{3j}, C_{4j} ...... C_{nj} with $n \ll N_m$.

11.2 Objectivity and the Undefined Past

According to our theory, the universe is completely determined by the way of the decomposition of the super quantum state Dao and thus by the accompanied consciousness. Therefore, it seems that the objectivity of the universe is lost. But it should be noted that Dao is objective and the universe is mostly determined by the original structure of Dao. When a consciousness is a consciousness, it is already there and it actually makes no choice. Certainly, the consciousness does not really determine the structure of universe.

By the current theory, the far past and the time before a consciousness is alive are actually not defined at all. It might indicates that all things happened before I was born were actually not real. It is true in my eyes but not in my parents' eyes. So the history can be constructed by gluing many consciousness' accompanied universes together thanks to the quasi-uniqueness of the Hamiltonian. But what about the universe when there is no consciousness in it? The universe without any consciousness is really not defined and meaningless in our theory. So it is also meaningless to talk about Big Bang in our theory. Nevertheless, one can always do the extrapolation using the Hamiltonian to obtain the further past without any consciousness.

12. Conclusions

In this article, we have developed a general theoretical framework for the study of consciousness in terms of quantum entanglement. By separating a subsystem from a quantum state, we are able to define both time and the consciousness in a relatively precise and manageable way. Relative distance and mass can be also defined in terms of entanglement entropies between physical entities separated from a pure quantum state. Two consciousnesses' situation is also discussed, which might indicates that the nowadays computers cannot be conscious. Relativity of intrinsic and extrinsic properties of consciousness is specially discussed. Based on this relativity and the new interpretation of mass, we postulate that consciousness has a negative mass but this bold hypothesis still needs to be justified by experiments.

It should be noted that the current theory is far from being perfect and there are still many unresolved and challenging problems within the theory. For example, why is background space of our universe is approximately three dimensional? Can we explain this in terms of relative distances? Can we generate a correct mass spectrum of elementary particles based on our new interpretation of mass? What is the mechanism behind the transformation of the intrinsic properties in W into induced extrinsic? Tackling these problems might take a whole life time. But on the other side challenge always means opportunities.

Acknowledgement

Author thanks NSFC (Grant No. 21104010) for financial support.

Reference

Barbour J. Relative-distance Machian theories. Nature 1974; 249: 2.
Barbour J. The end of time: the next revolution in physics. Oxford: Oxford University Press, 1999.
Barbour J. The deep and suggestive principles of Leibnizian philosophy. The Harvard Review of Philosophy 2003; 11: 45-58.
Barbour, J. The nature of time. 2009. arXiv:0903.3489.
Bohm D. A new theory of the relationship of mind to matter. Philosophical Psychology 1990; 3: 271-286.
Bohm D. Wholeness and the implicate order. Routledge, 1980.
Chalmers DJ. Facing up to the problem of consciousness. J Conscious Stud 1995;2:200-218.
De Sousa A. Towards an integrative theory of consciousness-part 1 (neurobiological and cognitive models). Mens sana monographs 2013; 11: 100-150.
De Sousa A. Towards an integrative theory of consciousness-part 2 (an anthology of various other models). Mens sana monographs 2013; 11: 151-209.
DeWitt BS. In: DeWitt BS, Graham N (Eds), The many-worlds interpretation of quantum mechanics. Princeton, NJ: Princeton Univ Press, 1973.
Everett H. 'Relative State' formulation of quantum mechanics. Rev Mod Phys 1957; 29:454-462.
Hales C. Dual aspect science. J of Conscious Stud 2009; 16: 30-73.
Hameroff SR. The brain is both neurocomputer and quantum computer. Cognitive Sci 2007; 31: 1035-1045.
Hameroff S. Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness. NeuroQuantology 2007; 5(1):1-8.
Hameroff SR and Penrose R. Conscious Events as Orchestrated Spacetime Selections. NeuroQuantology 2003; 1: 10-35.
Hameroff SR and Penrose R. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. Neural Network World 1995; 5: 793-804.
Joos E, Zeh HD, Kiefer C, Guilini D, Kupsch J and Stamatescu IO.Decoherence and the Appearance of a Classical World in Quantum Theory. Berlin: Springer, 2003.
Kiefer C. Quantum Gravity. Oxford: Oxford University Press, 2007.
Mensky MB. Mathematical Models of Subjective Preferences in Quantum Concept of Consciousness. NeuroQuantology 2011; 9: 614-620.
Mensky MB. Postcorrection and Mathematical Model of Life In Extended Everett's Concept. NeuroQuantology 2007; 5: 363-376.
Mensky MB. Everett interpretation and quantum concept of consciousness. NeuroQuantology 2013; 1: 85-96.
Smith CUM. The 'hard problem' and the quantum physicists. Part 1: The first generation. Brain Cogn 2006; 61: 181-188.
Smith CUM. The 'hard problem' and the quantum physicists. Part 2: Modern times. Brain Cogn 2009; 71: 54-63.