# Non-Computability of Consciousness

Daegene Song*

February 1, 2008

**Abstract**

With the great success in simulating many intelligent behaviors using computing devices, there has been an ongoing debate whether all conscious activities are computational processes. In this paper, the answer to this question is shown to be no. A certain phenomenon of consciousness is demonstrated to be fully represented as a computational process using a quantum computer. Based on the computability criterion discussed with Turing machines, the model constructed is shown to necessarily involve a non-computable element. The concept that this is solely a quantum effect and does not work for a classical case is also discussed.

## 1  Introduction.

Research in the field of artificial intelligence, which attempts to imitate and simulate intelligent activities using a machine, has blossomed along with the development of information technology [1]. Because the study of artificial intelligence has provided many insights into intelligent behaviors such as pattern recognition, decision theory, etc., there is a question whether consciousness or self-awareness could emerge out of a computational system, a view termed as strong artificial intelligence. This question can be rephrased and stated as follows: *Are all conscious activities computational processes?.* In this paper, the answer to this question is shown to be no.

In order to examine the computability of a physical phenomenon, the phenomenon should first be represented as a computational model; subsequently, the computability of this particular model can be examined. The physical phenomenon can then be claimed to be computable or not based on this examination. A similar approach will be taken in order to examine the computability of consciousness. Because consciousness is a phenomenon experienced by an observer, representation of consciousness as a computational process will be attempted and its computability will be examined. Although traditional approaches for studying consciousness have included neuroscience [2] or neural

---

*School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Korea.; Email: `dsong@kias.re.kr`.
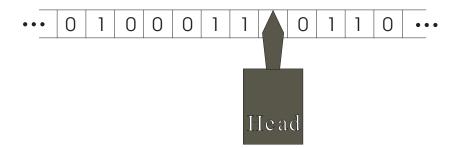
Figure 1: Turing machine. A Turing machine is an abstract model of a computing system consisting of internal states, a tape containing symbols in each cell and a head that reads and writes the symbol. Evolution in time of the Turing machine is described by $(I, a) \rightarrow (I', a', d)$ where $I$ is the internal state, and $a$ is a symbol written on the tape. At the $i$th cell, i.e., the head's position, the head reads the symbol $a$, and, with the instruction $I$, it writes a new symbol, $a'$, and moves either one cell to the left $(d = -1)$ or to the right $(d = +1)$ with an updated internal state, $I'$. The initiation and termination of computation are indicated by internal states, $h_0$ and $h_1$, respectively.

network modeling [3, 4], it is demonstrated that a quantum system to be presented below, it necessarily involves a conscious, as opposed to a physical, activity of an observer observing the unitary dynamics of a quantum state. Based on this observation, a particular quantum computer can be built such that it yields a computational model involving consciousness. Using logic similar to that in Turing's haling problem, it can be shown that this computational model necessarily runs into a contradiction. As a result, this effectively provides a counter-example to the assumption that all conscious activities are computational processes.

In this paper, it is not claimed that all conscious activities can be constructed, using a quantum computer, nor that they are quantum mechanical. Instead, it will be argued that the quantum system to be presented necessarily involves a certain conscious activity and that quantum theory provides a full description of this particular conscious activity. This argument will be used to build a quantum computing machine such that it suffices to provide a counter-example. A single counter-example is sufficient to prove the assumption is incorrect.

## 2 Computability and Turing machine.

In order to discuss computability of consciousness, let us first consider what it means to be computable. This can be done using the notion of Turing machines. A Turing machine, denoted as TM, is a theoretical model of dynamic computing system configured with an internal state, a tape containing a symbol in each cell

and identification of the position of the head on the tape (see Fig. 1). The time evolution of the TM is described by $(I, a) \rightarrow (I', a', d)$ where $I$ is the internal state, $a$ a symbol in the cell of the tape, and $d = \pm 1$. Therefore, at the $i$th cell, the head reads the symbol, $a$. Then, given the current internal state, $I$, which provides an instruction, the new symbol, $a' = I(a)$, is written, and the internal state is updated to $I'$ and the head moves either one cell to the right $(d = +1)$ or one cell to the left $(d = -1)$, i.e., to the $(i + d)$th cell.

Among the functions of the internal state, is the indication of initiation and termination of computation. Initially, this particular state is set to $h_0$, indicating the initiation of the computation. After the computation is completed, the state is set to $h_1$ and the machine ceases its activity. The output of the computation corresponds to the symbol in the cell where the head is located when the machine halts. For a given input, $i$, the TM runs, following the time evolution described above, and either (A) produces an outcome, $f = \text{TM}(i)$, and halts with the internal state set to $h_1$ or (B) loops forever and the internal state never reaches $h_1$. A system is called computable if it corresponds to a TM such that it follows either (A) or (B) for a given input $i$, and is called non-computable otherwise.

The issue of computability is considered in the following setup: suppose $\mathcal{T}$ is defined to have the following two properties:

1. Computable: $\mathcal{T}(i)$ when $i \neq \mathcal{T}$

2. Non-computable: $\mathcal{T}(i)$ when $i = \mathcal{T}$

That is, $\mathcal{T}$ corresponds to a TM that follows either (A) or (B) except when the input is $\mathcal{T}$, i.e., the description of $\mathcal{T}$, itself (see Fig. 2). In the following approach, the manner in which the computational model involving consciousness may be defined through $\mathcal{T}$ will be shown using a quantum computing machine, such that the two conditions regarding computability are satisfied, i.e. necessarily containing non-computability.

## 3    The halting problem.

Before proceeding with the discussion of consciousness, it is instructive to review Turing's halting problem [5] and to examine its use of a property similar to that seen in $\mathcal{T}$, such that the problem was shown to be non-computable. The situation for the halting problem is as follows: for some TMs, an outcome indicated by the halt internal state $h_0 \rightarrow h_1$, is computed, while for some other TMs, with a given input, the computation loops forever indicated by a constant internal state, $h_0$. Turing's halting problem asks if there is a TM that can distinguish between the two types given an arbitrary TM and an input. Suppose there is such a TM that performs a calculation given the description of TM and $i$ such that it is able to determine if TM halts or not. This assumption then makes it possible to construct a particular TM, $\text{TM}_\text{H}$, such that the machine does not halt, for an input TM, if and only if TM(TM) halts. However, a contradiction follows for $\text{TM}_\text{H}$ when the input is $\text{TM}_\text{H}$ itself, because $\text{TM}_\text{H}(\text{TM}_\text{H})$ does not halt, if and only if, $\text{TM}_\text{H}(\text{TM}_\text{H})$ halts.
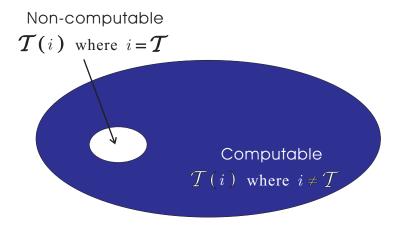
Figure 2: Computability and Non-computability. $\mathcal{T}$ is defined to have the property corresponding to a Turing machine that either halts or not unless it is given an input of $\mathcal{T}$ itself. The halting problem can be defined in association with $\mathcal{T}$, such that it necessarily is non-computable. Similarly, quantum theory allows consciousness to be represented as a computational process in terms of $\mathcal{T}$, such that it would necessarily consist of a non-computable element when the input is $\mathcal{T}$ itself.

Therefore, by identifying the TM$_H$ associated with the hypothetical TM that could decide if an arbitrary TM would halt on a given input, it is possible to show that the TM$_H$ contains an element that neither halts after completing the computation nor loops forever as should a valid TM. The constructed TM$_H$ has the same property as $\mathcal{T}$ in Fig. 2, i.e., it is computable except when the input is TM$_H$ itself.

## 4 Conscious activity in quantum system.

In order to represent a phenomenon of consciousness as a computational model, the manner in which a conscious activity is involved in a quantum system is first discussed. This will be conducted using the notation of a qubit, a two-level quantum system. A qubit in a density matrix form is written as $|\psi\rangle\langle\psi| = \frac{1}{2}(\mathbf{1} + \hat{\mu} \cdot \vec{\sigma})$ where $\hat{\mu} = (\mu_x, \mu_y, \mu_z) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ with $\sigma_x = |0\rangle\langle1| + |1\rangle\langle0|$, $\sigma_y = -i|0\rangle\langle1| + i|1\rangle\langle0|$, and $\sigma_z = |0\rangle\langle0| - |1\rangle\langle1|$. Therefore a qubit, $|\psi\rangle\langle\psi|$, can be represented as a unit vector $\hat{\mu} = (\mu_x, \mu_y, \mu_z)$ pointing in $(\theta, \phi)$ of a sphere with $0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi$. In quantum theory, there is another important variable called an observable. For a single qubit, an observable can also be written as a unit vector, $\hat{\nu} = (\nu_x, \nu_y, \nu_z) = (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta)$, pointing $(\vartheta, \varphi)$ direction in a sphere. Therefore if one is to make a measurement in $(\vartheta, \varphi)$ direction, the observable would be $\hat{\nu} \cdot \vec{\sigma}$.

Representing a qubit and an observable as unit vectors in the Bloch sphere will make their visualization easier which will be helpful in the following discussions.

Let us consider one particular phenomenon, denoted as **P1**, and described as follows: *an observer observes the unitary evolution of a qubit, $\hat{\mu}$, with respect to the observable, $\hat{\nu}$.* The observer is observing the evolution of $\hat{\mu}$ indirectly and a measurement can be followed in order to confirm the evolution. When a measurement on $\hat{\mu}$, with the observable, $\hat{\nu}$, is made, it yields a real eigenvalue that can be directly observed by the observer. Before discussing the description of the phenomenon, **P1**, using the dynamics of quantum theory, it is necessary to illustrate why the phenomenon, **P1**, necessarily involves a conscious activity of the observer. An observable serves as a coordinate or a reference frame when the measurement is made on a given state vector [6]. This concept is easier to visualize with two unit vectors, $\hat{\mu}$ and $\hat{\nu}$. The unit vector representing an observable, i.e., $\hat{\nu}$, is serving the role of a coordinate for the unit vector representing a qubit, $\hat{\mu}$. Because the measurement is performed by an observer, the observable is considered to be a coordinate or a reference frame of the observer, for a given qubit $\hat{\mu}$.

However, in quantum theory, observables, being a reference frame of the observer, are fundamentally different from reference frames in classical physics. In quantum theory, the state vectors have representation in, and evolve in, the Hilbert space, a complex vector space. This description was invented in order to correctly predict the outcome of measurement performed on a state vector which yields a real eigenvalue outcome. Not only do state vectors reside and evolve in the Hilbert space, so do the observables. Because the observables correspond to the reference frame of the observer and they exist in the complex Hilbert space, it must be concluded that, unlike reference frames in classical physics, quantum observables correspond to an observer's reference frame in thought. That is, an observable should be considered to represent the conscious status of an observer while observing a given state vector. This argument explains why the phenomenon, **P1**, necessarily involves a conscious activity.

The qubit in **P1** can be any 2-level quantum system, for example, a spin 1/2-particle or any quantum system in 2-levels, etc. However, it is not necessary to specify all properties of the physical system other than $\hat{\mu}$, because $\hat{\mu}$ is a pure state and is disentangled from the state that represents other properties of that quantum system. Therefore, as far as the phenomenon, **P1**, is concerned, $\hat{\mu}$ provides a full description of the physical system. The same logic applies to $\hat{\nu}$ as well. The vector, $\hat{\nu}$, is not entangled with vectors representing other observables. Therefore, $\hat{\nu}$ must provide a full description of the conscious status of the observer in phenomenon, **P1**. That is, similarly to the case with $\hat{\mu}$, it is not necessary to be concerned with other conscious activities of the observer because $\hat{\nu}$ is disentangled from them. Therefore, **P1** not only necessarily involves a conscious activity of the observer, $\hat{\nu}$ gives a full description of the conscious activity as far as the phenomenon, **P1**, is concerned.

Quantum theory provides two approaches in describing the natural phenomenon **P1**. Given $\hat{\mu}$ and $\hat{\nu}$, the first is by applying a unitary operation to the qubit with $\hat{\mu}' = U\hat{\mu}U^{\dagger}$ where a measurement would yield the expectation

value of $\hat{\nu} \cdot \hat{\mu}'$. The second is by applying a unitary operation to the observable as $\hat{\nu}' = U^{\dagger} \hat{\nu} U$ and a measurement would yield $\hat{\nu}' \cdot \hat{\mu}$. That is, quantum theory insists that, in order to have an observer observe the unitary transformation of $\hat{\mu}$ with respect to $\hat{\nu}$, either a unitary transformation is applied to the qubit, i.e., the first approach, or the observer's reference frame $\hat{\nu}$ is changed, i.e., the second approach. The first approach is called the Schrödinger picture and the second corresponds to the Heisenberg picture.

In the second approach, it was the observable that went through a unitary transformation which should describe the same phenomenon, **P1**, as the first approach. Because the evolution of observables through unitary transformations are performed in the Hilbert space and the observable is the observer's conscious status in **P1**, an observable that is being changed must correspond to a conscious activity of an observer. However, while the observer's conscious status is being changed, the observer is not observing the observable but the state vector, $\hat{\mu}$. Therefore, this approach also yields the description of the natural phenomenon, **P1**, just as in the first approach.

# 5 Conscious activity in quantum computing process.

So far, it has been argued that the phenomenon, **P1**, necessarily involves an observer's conscious activity and quantum theory provides a full description of the conscious status of the observer regarding **P1**. Based on these observations, a quantum computational model is to be constructed such that it represents a phenomenon involving a conscious activity and its computability will be examined. In particular, $\mathcal{T}$ will be defined in terms of this computational model and the computability for a given input, $i$, will be examined. In the next section, it will be argued that when the input is $\mathcal{T}$ itself, it represents consciousness and will be proven to be non-computable similarly to the halting problem.

Let us review basic elements of quantum computation by following the discussion in [7, 8]. The particular class of quantum computers to be considered is assumed to perform a computation on an input of a single qubit, i.e., a unit vector in the Bloch sphere, $\hat{\mu}_s$, in which the subscript, $s$, is placed in order to distinguish it from the halt qubit to be defined shortly. In this particular class, the computation is assumed to be conducted through a unitary process on a given single input qubit, a rotation about the $y$-axis by $\delta$, i.e., $U_y \equiv \cos \frac{\delta}{2} |0\rangle\langle 0| - \sin \frac{\delta}{2} |0\rangle\langle 1| + \sin \frac{\delta}{2} |1\rangle\langle 0| + \cos \frac{\delta}{2} |1\rangle\langle 1|$. Among the components of the classical TM, the head exists which reads each cell on the tape (see Fig. 1). The head in the classical TM may correspond to the observables in the quantum computer. As demonstrated earlier, for a single qubit, the observable can also be written as a unit vector in the Bloch sphere, which will be denoted as $\hat{\nu}_s$.

As suggested in [8], in addition to the system input qubit, an additional qubit is placed which indicates if the computation on the system qubit has successfully

ended by $0 \rightarrow 1$ after a valid computation on the single system qubit which remains 0 otherwise. This is equivalent to the classical TM in which its internal state indicates if the machine completed its computation by $h_0 \rightarrow h_1$. The halt qubit is set to point at the $z$-direction, i.e. $\hat{\mu}_h = (0, 0, 1)$. The corresponding observable, $\hat{\nu}_h = (0, 0, 1)$, also set to point at the $z$-direction, initially. Therefore, the quantum computer constructed for a given input $\hat{\mu}_s$, is a closed system consisting of $\hat{\mu}_s, \hat{\nu}_s, \hat{\mu}_h$, and $\hat{\nu}_h$. Because there is freedom to set the observable, it can be used to identify one particular quantum computer which works on a given input, $\hat{\mu}_s$. Among the infinitely many choices of $\hat{\nu}_s$, assume that one particular quantum computer exists with the observable, $\hat{\nu}_s = (0, 0, 1)$. Because the unitary evolution will be $U_y$ only, the initial observable fully characterizes this particular quantum computer and it will be defined as $\mathcal{T}$.

The quantum model constructed operates on a single qubit, and, only a single operation, i.e., $U_y$, is considered. Therefore, there is no need to specify any internal state that yields an instruction because there is only one operation. The only internal state needed is the indication of initiation and termination of the computation that is represented with the halt qubit. Moreover, indication of the position of the head is unnecessary because there is only one qubit, which corresponds to a tape with a single cell. Therefore, the quantum computer constructed corresponds to a very simple case of a quantum mechanical TM.

One particular phenomenon, denoted as **P2**, is considered as follows: *an observer observes a rotation of the input, $\hat{\mu}_s$, about the y-axis by $\delta$, with respect to $\hat{\nu}_s$.* As in **P1**, the observer is observing the rotation indirectly and a measurement on $\hat{\mu}_s$ with the observable, $\hat{\nu}_s$, can be followed to confirm the evolution. Note that **P2** is almost identical to the phenomenon, **P1**, except the unitary operation is specified as $U_y$. Therefore, similarly to the case with **P1**, the phenomenon, **P2**, necessarily involves a conscious activity of the observer, represented as $\hat{\nu}_s$. Moreover, as discussed with the instance of **P1**, $\hat{\nu}_s$ provides a full description of the conscious status of the observer in reference to **P2**. In the following, it will be established that the quantum computer constructed, $\mathcal{T}$, represents **P2** as a computational model and is computable, therefore indicating that the phenomenon, **P2**, is computable.

As discussed earlier, the quantum theory provides two approaches for the evolution in time of a quantum system. Therefore, because $\mathcal{T}$, the quantum computer constructed, is a quantum system, it should also evolve in both approaches. The evolution in time of $\mathcal{T}$ with an initial input state, $i = \hat{\mu}_s = (0, 0, 1)$, will be examined. The first approach, i.e., the Schrödinger picture, is considered as follows: the unitary operation $U_y$ transforms the input as $\hat{\mu}_s \rightarrow U_y \hat{\mu}_s U_y^\dagger$ and the halt qubit $\hat{\mu}_h \equiv (0, 0, 1)$ halts by transforming into $-\hat{\mu}_h$. In the second approach, i.e., the Heisenberg picture, it is the observable that evolves. Therefore, $U_y^\dagger$ transforms the vector representing the observable $\hat{\nu}_s$ into $U_y^\dagger \hat{\nu}_s U_y$ and the observable for the halt qubit $\hat{\nu}_h \equiv (0, 0, 1)$ is transformed into $-\hat{\nu}_h$. Therefore, in the second approach, the observer's conscious status $\hat{\nu}_s$ is being changed while the observer observes $\hat{\mu}_s$. This should yield the same observation as the first approach. It is noted that the expectation

7

value of $(U_y^\dagger \hat{\nu}_s U_y) \cdot \hat{\mu}_s$ for the second approach is equal to the expectation value in the first approach, $\hat{\nu}_s \cdot (U_y \hat{\mu}_s U_y^\dagger)$. Therefore, both the first and the second computational processes ultimately describe the phenomenon, **P2**, by correctly producing an outcome described in **P2**.

Initially, it was discussed that a system is stated to be computable when it satisfies one of two criteria, i.e. either (A) it halts after completion of a valid computation or (B) it loops forever without halting. $\mathcal{T}$ was shown to yield the description of **P2**, with a given input $\hat{\mu}_s$, by following both pictures in quantum theory, i.e., both approaches yielded the outcome by which $\hat{\mu}_s$ rotated about the $y$-axis by $\delta$, with respect to $\hat{\nu}_s$, and halted. Therefore, the phenomenon, **P2**, can be claimed to be computable because its computational representation, $\mathcal{T}$, with the input $\hat{\mu}_s$, was shown to be computable by satisfying the criterion (A).

# 6   Counter-Example to the Assumption.

In case of the Heisenberg picture description of **P2**, as well as of **P1**, it was discussed that the observer is in the conscious status undergoing change, $\hat{\nu}_s$, and observes $\hat{\mu}_s$. This was shown to yield the phenomenon of **P2**, i.e., the observer observing the rotation of $\hat{\mu}_s$. A slightly different case can be considered. While the observer is in the conscious status, $\hat{\nu}_s$, that is being changed, the observer observes $\hat{\nu}_s$ rather than $\hat{\mu}_s$. This is a peculiar aspect of consciousness–observing one's own conscious status–that is not observed in other measurement experiences, for example, in classical dynamics. This phenomenon can be stated as follows and denoted as **P3**: *an observer observes a rotation of the input, $\hat{\nu}_s$, about the $y$-axis by $\delta$, with respect to $\hat{\nu}_s$.* Therefore, in **P3** which describes consciousness of the observer, $\hat{\nu}_s$ is serving the role of a state vector, because it is being observed, and an observable, because it is serving as the reference frame of the observer. Unlike the cases of **P1** and **P2**, the measurement confirmation is not needed for **P3**. While the conscious status, $\hat{\nu}_s$, is evolving, the observer is not observing $\hat{\mu}_s$ but $\hat{\nu}_s$. No measurement is needed in order to confirm the evolution of $\hat{\nu}_s$ because the observer is already experiencing it as consciousness.

In the previous section, it was demonstrated that $\mathcal{T}$, with an input $\hat{\mu}_s$, provides a computational model for describing the phenomenon of **P2** and was shown to be computable. Because **P3** is exactly the same as **P2** except the input has changed to vector, $\hat{\nu}_s$, from $\hat{\mu}_s$, it follows that $\mathcal{T}$, with an input, $\hat{\nu}_s$, must correspond to a computational model representing the phenomenon, **P3** (see Table 1). The observable, $\hat{\nu}_s = (0, 0, 1)$, fully characterizes $\mathcal{T}$. Therefore, $\mathcal{T}$ with an input, $\hat{\nu}_s$, can also be stated as $\mathcal{T}$ with an input of the description of $\mathcal{T}$, or simply as $\mathcal{T}$ with an input, $\mathcal{T}$. In the following, the computability of $\mathcal{T}$ for a given input of $\mathcal{T}$, which represents the phenomenon, **P3**, as a computational model, is to be examined.

As established previously, quantum theory provides two approaches to the evolution in time of $\mathcal{T}$ for the input, $\hat{\nu}_s$, because it is a quantum system where $\hat{\nu}_s$ corresponds to both a state and an observable. In the first approach, it is the input system that evolves. Since the input is $\hat{\nu}_s$, the evolution is as follows,

| Computational Model | Phenomenon |
|---|---|
| $(\mathcal{T}, i = \hat{\mu}_s)$ | **P2**: Observer observes the rotation of $\hat{\mu}_s$ with respect to $\hat{\nu}_s$. |
| $(\mathcal{T}, i = \hat{\nu}_s)$ | **P3**: Observer observes the rotation of $\hat{\nu}_s$ with respect to $\hat{\nu}_s$. |

Table 1: Analogy between the computational model, $\mathcal{T}$, and phenomena **P2** and **P3**. If the phenomenon, **P2**, can be represented as a computational model, $\mathcal{T}$, with an input, $\hat{\mu}_s$, then $\mathcal{T}$ with an input, $\hat{\nu}_s$, should correspond to a computational model for the phenomenon, **P3**.

$\hat{\nu}_s \rightarrow \hat{\nu}'_s = (\sin\delta, 0, \cos\delta)$, while the halt qubit is transformed as $\hat{\mu}_h \rightarrow -\hat{\mu}_h$. Quantum theory provides a second approach where the same vector, being an observable, is transformed as $\hat{\nu}_s \rightarrow \hat{\nu}''_s = (-\sin\delta, 0, \cos\delta)$, while the observable for the halt qubit is transformed as $\hat{\nu}_h \rightarrow -\hat{\nu}_h$. It is noted that $\hat{\nu}'_s \neq \hat{\nu}''_s$ unless $\delta = k\pi$ where $k = 0, 1, 2, ....$

Let us now discuss the computability of $\mathcal{T}(i)$ where $i = \mathcal{T}$. In order for $\mathcal{T}(\mathcal{T})$ to be computable, it has to follow either the computability criterion (A) or (B). Since $\mathcal{T}$ halted on both approaches, i.e., $\hat{\mu}_h \rightarrow -\hat{\mu}_h$, with respect to $\hat{\nu}_h$, in both pictures, it must follow (A) rather than (B) in order to be computable. In order to satisfy (A), the halt qubit of $\mathcal{T}$ must have halted accompanied by a valid computation, i.e., both approaches should yield the same outcome predicted in **P3**. However, the two approaches yielded two generally different outcomes for a single input vector, $\hat{\nu}_s$. The second approach did not yield the outcome described in the phenomenon, **P3**, because the vector is rotated by $-\delta$. Therefore, this results in a contradiction because $\mathcal{T}$ halted on the invalid computation. The contradiction is noted to result from a peculiar property of consciousness in which $\hat{\nu}_s$ is serving as a reference frame of the observer and as a system to be observed.

The assumption states that all conscious activities are computational processes. Because $\mathcal{T}(i)$, with $i = \hat{\nu}_s$, being a computational model of the phenomenon **P3**, is a closed and independent system, this must satisfy the assumption. However, it was shown that $\mathcal{T}$, with a given input $\hat{\nu}_s$, is not computable. That is, a particular conscious activity of an observer observing the change of an observable, as described in **P3**, is not computable. Therefore, this leads to a conclusion that the assumption is incorrect, because it suffices to have a single counter-example to invalidate the assumption.

Perhaps, by considering a larger system that includes the qubit, the contradiction may be removed and may yield the result that consciousness is always a computational process. This is commonly seen in thermodynamics in which a subsystem violates the second law but this violation is always removed when the total system is considered. However, this kind of argument would not work because the evolution considered in **P2** and **P3** are for pure states. Any attachment of ancilla to $\mathcal{T}$ and their interaction with the system qubit would cause entanglement and this will not properly represent the physical phenomena **P2** and **P3**.

# 7 Discussion.

The above argument applies only as a quantum effect. The classical TM cannot define consciousness using the same technique. As discussed, a reference frame of quantum measurement was represented in complex Hilbert space which led to the conclusion that it must correspond to the observer's conscious status. A classical measurement yields an outcome in terms of the difference between the object and the reference frame of an observer, and, unlike consciousness, the observer cannot observe the dynamics of its own reference frame alone. Therefore, the same argument used with the quantum computing machine involving conscious activities cannot be used in a classical case.

In [9], Penrose discussed that a non-computable aspect in consciousness may exist at the fundamental level as described in Gödel's incompleteness theorem. Including Turing's halting problem, there have been a number of mathematical examples showing undecidability in Gödel's theorem. In this paper, it was demonstrated that, as in Penrose's suggestion, consciousness is a physical, i.e., rather than mathematical, example of Gödel-type proofs.

# References

[1] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, 2nd edition, 2002)

[2] C. Koch, *The Quest for Consciousness: a Neurobiological Approach* (Roberts & Company Publishers, 2004).

[3] G. Tononi, G.M. Edelman, *Science* **282**, 1846 (1998).

[4] R.L. Harvey, *Neural Network Principles* (Prentice-Hall, Englewood Cliffs, NJ, 1994).

[5] A.M. Turing, *Proc. London Math. Soc.* (2) **442**, 230 (1936).

[6] A. Peres, *Quantum Theory*, Kluwer Academic Publishers, (1991).

[7] P.A. Benioff, *J. Stat. Phys.*, 563 **22** (1980).

[8] D. Deutsch, *Proc. R. Soc. London A* **400**, 97 (1985).

[9] R. Penrose, *The Emperer's New Mind* (Oxford University Press, New York, 1989).