

# Consciousness and self-location

Yu Feng  
foundationsofqm@gmail.com

October 4, 2023

## 1 Two realities

The philosophical investigation of quantum mechanics may shed some light to the mysterious nature of consciousness. According to the many-world interpretation, there are many versions of me perceiving different versions of reality. In particular, I would call the perceived reality (or perceived world) “local reality”, while the subject-agnostic reality (e.g. the cosmic wave function) “global reality”. It is natural to relate consciousness to the “localisation” of realities. In this paper, I will illustrate this idea and show that such a relation is plausible.

For clarity, I will define some terms first. The term consciousness refers to subjective experience. It is not necessarily the kind of subjective experience that a human being, or even any living organs would experience. In general, consciousness denotes the phenomenal property, which is the property of what it is like to be something. Apart from phenomenal properties, there are also physical properties, which characterise something from outside. Formally, (global) physical properties of an object A are its characteristics irrespective of any particular observer. Such properties can be described by either

- a) A is in ... (e.g. A is in a state of reflecting all blue light) with probability of ..., or
- b) If A is in ..., then ... (e.g. if A is pushed by a force F, then it will accelerate at a rate denoted by a. This property gives A’s mass  $m=F/a$ ).

Now suppose Everett is right, and the entire cosmos is in a giant superposed quantum state. The state of cosmos follows unitary evolution and is completely deterministic. However, this entirety is not the world we actually perceive. In fact, due to its superposed structure, it can't be perceived at all. Just consider a mini-cosmos consisting of only me and Schrödinger's cat. The entirety is in a superposition of two states,  $|\text{living cat, me perceiving living cat}\rangle$  and  $|\text{dead cat, me perceiving dead cat}\rangle$ . Our actual perception is limited to either one of the two branches. Therefore, the reality we perceive is not the reality of the entirety. It is "localised" to a particular version of me. We may then use "local" and "global" to differentiate the two types of realities.

## 2 Entropies and self location

In [1], the author shows that how the two types of realities lead to two types of entropies. The global reality of the entirety, which is assumed to be in a pure state, has a constant entropy of zero. This is not true for local realities. For instance, if the mini-cosmos in the previous paragraph is in a superposed state of  $(|\text{living cat, me perceiving living cat}\rangle + |\text{dead cat, me perceiving dead cat}\rangle)/\sqrt{2}$ , then the local realities I might perceive are (living cat, me perceiving living cat) and (dead cat, me perceiving dead cat), each of which has a 50% chance. This leads to a non-zero entropy of 1 bit. We may regard it as the result of the uncertainty about my self location: it is equally probable for me to actually live in either of the two local realities. Therefore, this entropy measures the uncertainty due to the absence of self-locating knowledge.

This entropy is coined as the objective entropy. The name comes from the fact that the entropy is independent or agnostic of any subject or subjective experience. One can regard the objective entropy as the entropy from the perspective of a third person, or an observer outside the whole system. However, practically any observer has some interaction with the observed. As a result, the observer's subjective experience always plays a role in a measurement. In the previous example, I would not agree with the 1-bit entropy about the cat's status, for there is no uncertainty at all. I should either perceive a living cat, or perceive a dead cat, without ambiguity. My subjective experience "reduces" the global reality to a particular local reality. Just like picking one ball from two identical balls, this "reduction" carries 1-bit Shannon information. As a result, my subjective experience results in a 1-bit decrease in entropy. This entropy reduction due to knowing one's own self location is coined as the subjective entropy, which partially or fully offsets the objective entropy.

Regarding entropy as a measure of lack of information, we may now consider the completeness of physical properties defined in a) and b). For the entire cosmos, the entropy is zero indicating no lack of information at all. If a fictitious observer outside the entirety knows all its physical properties, or equivalently identifies its quantum state, she would know everything about it. This means the entirety is completely characterised by its physical properties. But if she divides the entirety into two entangled subsystems, A and B, then non-zero entropy arises:  $S(A) = S(B) > 0$ . The physical properties of A can at most tell as much as what the (mixed) quantum state of A can tell. Its entropy  $S(A) > 0$  implies that the physical properties of the subsystem are incomplete.

How can the physical properties become incomplete for partitions of the entirety? In the previous example of Schrödinger's cat, the global reality contains two local realities in parallel. In the first local reality the cat is alive and I perceive a living cat; in the second local reality, the cat is dead and I perceive a dead cat. The state of the global reality is a superposition of the states of the two local realities. Due to the linearity of quantum mechanics, the two local realities evolve independently according to the physical properties. But as they are independent of any particular subject, these physical properties cannot tell which of the two local realities I actually live in. For an observer outside the system, the two local realities evolve in parallel and there is no need to specify one over the other. But for the observer, I, which is inside the system, ambiguity arises as what I observe does depend on which local reality I actually live. With the presence of subjective experience, such ambiguity is removed by the fact that I know my self location from direct experience.

### 3 Subject-object division

In general, we may consider the bi-partition of the entirety. We call it the subject-object division, though the assignment of either subsystem to the subject is discretionary. This means that the entirety is a bipartite system consisting of a subject I and all the rest as the object O. Since physical properties do not account for any subjective experience, they can at most provide a probabilistic description about O, e.g. I could either live in a world where the cat is alive or live in a world where the cat is dead. The lack of information in the physical properties, measured by a positive objective entropy, is erased by the subjective experience of I. By specifying the self location, it gives a negative subjective entropy which offsets the positive objective entropy.

This implies that the properties about the subject's self location, or namely self-

locating properties, complement the physical properties. To illustrate this point rigorously, consider a partition of the entirety into A and B. The entirety is in a state of  $\sum c_i |a_i b_i\rangle$ , where each branch corresponds to a local reality which may be perceived by A or B. The physical properties defined here are agnostic about their self locations, and thus characterise the subsystems probabilistically. This uncertainty is measured by a positive objective entropy  $S(A) = S(B) = -\sum_i |c_i|^2 \log |c_i|^2$ . The self-locating properties, however, remove such uncertainty by pointing out which local reality A or B actually perceives. The subjective entropy measures the information gain given by the self-locating properties of subsystem A or B,  $S'(A) = S'(B) = \sum_i |c_i|^2 \log |c_i|^2$ .  $S'(A)$  offsets  $S(A)$  and  $S'(B)$  offsets  $S(B)$  exactly. From the perspective of both A and B, they live in a deterministic world (as far as their subjective experience may reach). The total entropy being zero indicates that self-locating properties complete the informational gap left by physical properties.

## 4 The role of subjective experience

We have seen that under system partitions, e.g. the subject-object division, physical properties are insufficient characteristics of the local reality. The complementary part is one's self location provided by subjective experience. This naturally leads to a version of panpsychism, which suggests that any part of the entirety has some sort of subjective experience. The subjective experience serves to identify one's self location among quantum-mechanical branches. Put our argument formally,

1. Any subsystem is associated with some self-locating properties for a meaningful description of the local reality it inhabits.
2. The self-locating properties are grounded in subjective experience.
3. Any subsystem is associated with some subjective experience.

In this argument, point 2 follows the intuition that it is the subjective experience that provides the information about which branch of the cosmic wave function one inhabits. However, since our subjective experience is so limited, this posterior belief may not apply to other subsystems. It is as well possible that the self-locating properties of an electron are fundamental, or grounded in properties other than subjective experience. In order to generalize our own phenomenal properties to other subsystems, we formulate the argument below to show that self locations of fundamental particles are indeed grounded in (pro)phenomenal properties:

1. My own self-locating properties are wholly grounded in (macro)phenomenal properties.
2. My own self-locating properties are wholly grounded in self-locating properties of my constituents, e.g. fundamental particles that form my body.
3. If my self-locating properties are wholly grounded in both my phenomenal properties and my constituents' self-locating properties, then either one constitutes the other, or one is grounded in some properties which constitute the other.
4. If my constituents' self-locating properties constitute my own phenomenal properties, then they themselves are prophenomenal properties.
5. If my constituents' self-locating properties are grounded in some properties that constitute my phenomenal properties, then the grounding properties are prophenomenal properties.
6. In either case, the self-locating properties of my constituent particles are grounded in some prophenomenal properties.

From our analysis, partition of the entirety entails the presence of subjective experience. We may thus view consciousness as a sequence of such partition, in particular, the subject-object division. From a holistic point of view, subject-object division gives rise to the apparent separation of subjective experience and objective physical properties, even though they are one from the perspective of the entirety. From a functional point of view, physical properties and subjective properties play complementary roles for a partition of the entirety. Physical properties characterise the states and dynamics of each subsystem, and are generally probabilistic. Subjective properties thus provide a complementary function, by pointing out the self location among all quantum-mechanical branches. This ensures that our subjectively perceived world is always unambiguous.

## **5 About the nature of consciousness**

### **5.1 Causal closure**

Now we will examine some arguments about the mysterious nature of consciousness. The causality argument that leads to the epiphenomenal account of consciousness is given by the following argument:

1. Physical world is causally closed.
2. The physical world cannot be causally closed, if it is causally impacted by consciousness.
3. The physical world cannot be causally impacted by consciousness.

The fallacy is in point 2, which mixes different types of realities. The global reality is indeed causally closed. Its evolution is described by the Schrödinger's equation (or its partially traced-out version for a subsystem of the entirety). But the local reality, which is the reality being experienced by a subject, can actually be causally affected by subjective experience. When we consciously make a choice, e.g. open or not open the box containing Schrödinger's cat, we know that it will affect the outer world which we experience. The function of consciousness is not only passively experiencing a local reality, but also to a certain degree actively choosing which local reality to experience. In other words, the causal closeness of the physical world only applies to physical properties, which are in general probabilistic. Such uncertainty, measured by a positive objective entropy, leaves a gap in the causal closure. Within this gap, the local reality a subject experiences is not causally closed. This does not contradict the causality of physical laws, but instead complements it.

## 5.2 Epistemic gap

As discussed above, the epistemic gap between something's objective description and subjective experience cannot be bridged easily. One is causally closed while the other is not. One can be fully conveyed to a third person while the other can't. This is why the zombie argument and the knowledge argument are so frustrating. The subjective experience, which is what it is like to be something, can only be perceived from a first-person perspective. If it can be fully conveyed to a third person, then such communication may be regarded as an objective description of such experience. The subjective experience may then be regarded as part of the physical properties. However, as shown in this work, the collection of all physical properties can still not provide a sufficiently complete description of the subsystem. This means that the subjective experience has to be privileged, in the sense that it can only be comprehended by experiencing it as first person. Put it formally,

1. If subjective experience of a subsystem (first person) can be fully conveyed to another subsystem (third person), then there is an objective description of such experience.

2. If subjective experience has objective description, then it is part of the physical properties.
3. The collection of all physical properties and subjective experience provides a complete description of the subsystem.
4. If subjective experience is part of physical properties, the collection of all physical properties provides a complete description of the subsystem.
5. The collection of all physical properties cannot provide a complete description of the subsystem.
6. Subjective experience of a subsystem cannot be fully conveyed to a third person.

### **5.3 The unmanifest and its localisation to particular subject**

It is worthy mentioning that the physical properties discussed above are agnostic of any particular subject. This means that such physical properties are almost always probabilistic due to the nature of quantum mechanics. They therefore are not identical to the deterministic physical properties in the local reality where we live. In fact, the deterministic physical properties are the physical properties "localised" by the subjective experience. In the absence of subjective experience, the entirety does not manifest itself as a specific form, thus in an unmanifested state. Consciousness, however, shines light to this unmanifested state. It manifests the entirety as a particular form, which is the universe we perceive.

Essentially, the local reality is the reality localised by specific subjective experience. What is the difference between my local reality and yours? Well, any localisation requires some center, and here the center is "I", the sense of self. It is the sense of self that integrates the subjective experience associated with this center, both inside and outside of me. There is nothing special about this center, and one would expect the center of localisation is everywhere: every human being has a sense of self, every living organ probably has as well, and perhaps every particle has such center. This is in agreement with the previous analysis: any partition that separates a subsystem from the entirety gives rise to subjective experience. Such subjective experience is associated with this subsystem, which is the center for localisation and internally it may experience a sense of self.

As discussed above, the local reality is manifested as the entirety is localised according to a particular center, or a particular subsystem. Even for physically

identical subsystems, they are two different centers of localisation. It is therefore possible for them to experience different subjective experience. This means that inverted spectrum is possible (and logically conceivable). A robot that functions the same as me, in particular, does not share my inner experience. This is because the inner experience is not just about the outer functions. It is dependent of the substrate from which the functions are performed. This will become more clear in the following analysis that associates the subjective experience of a subsystem with its components.

## 5.4 Combination of subject and subjective experience

Note that by working from a top-down approach starting from the entirety, we essentially avoid the combination problem which lies in the way towards a panpsychist account of consciousness. But actually, combined consciousness is plausible from a many-world interpretation of quantum mechanics. Consider Wigner's friend who performs a measurement about Schrödinger's cat. The observer and the observed are in a state superposed by  $|\text{living cat, Wigner's friend perceiving living cat}\rangle$  and  $|\text{dead cat, Wigner's friend perceiving dead cat}\rangle$ . Now no matter Wigner observes the cat or not, Wigner and his friend, combined as one system or one team, would know if the cat is alive or dead. If none of them observes the cat, then the combined system does not possess any self-locating properties about the cat. It is therefore legitimate to state that the self-locating properties of a system are grounded in such properties of its components.

Since subjective experience gives rise to the self location, it is plausible to state that the subjective experience of a system is constituted by the subjective experience of its components. One may argue that individual minds do not aggregate into a compound mind, and the combined system of Wigner and his friend does not possess any subjective experience. But it is possible that subjective experience does exist for any combined system. It is just not received or processed by the individual brain. We all have some experience of alternative states of consciousness, from which unusual subjective experience is perceived. It is likely that such experience always exists, but is filtered by the brain under the normal state of consciousness.

From the author's own experience, our subjective experience, or macrophenomenal properties, is constituted by some primordial subjective experience, or prophenomenal properties. In eastern philosophy, such primordial experience is termed as qi, ki or prana. The characteristics of the primordial experience leads to the classification of fundamental elements (which are essentially fundamental elements of the phenomenal world) in many ancient philosophies, e.g. earth, water, fire and air. Under

certain combinations and/or motions, they give rise to everyday experience such as pain, fear, joy, excitement, etc.

## 6 Conclusion and implications

Overall, the entirety, uniting all subjects and objects, is complete and does not possess any particular physical properties or subjective experience. It is the superposition of all possibilities evolving unitarily. Its partition, however, gives rise to physical properties and subjective experience simultaneously. They play complementary roles to each other.

This gives some insight to the existential question: why there is something rather than nothing. Nothingness, or formlessness, is merely an approximate verbal description of the absolute entirety. In fact, no characteristics are suitable for the absolute, either in the form of physical properties or in the form of subjective experience. Then, the subject-object division comes in, forming a particular world (or local reality). It is always relative, with respect to certain subject, and is manifested as certain forms. These forms can be classified into two categories, physical properties and subjective experience. The latter cannot be conveyed to a third person, and cannot be reduced to the former. It in fact fills the informational gap which is impossible for the former to fill. The casual effect and the privilege of subjective experience are related to the role it plays, which is to identify the subject's self location.

## References

- [1] Feng, Y. Everettian Formulation of the Second Law of Thermodynamics. <https://philpapers.org/archive/FEND.pdf>