Properties of Consciousness, Qualia and Self-Reflection by Intentional Paths

by Sven Nilsen, 2017

Usually one consider consciousness as a philosophical domain with little relevance to real world problem solving, but this is no longer the case. The rapid advancement in the field of artificial intelligence requires these philosophical problems to be resolved before we have the technology to do irreversible changes to human consciousness. In particular, we want to make predictions about what changes in experiences can be anticipated when modifying brains. In this paper I show some properties of consciousness, qualia and self-reflection when modeling these phenomena as intentional paths.

Consider the following thought experiment:

Alice and Bob both experience the world from inside their own brains. One day Alice invents a device called the "Neuromod" which allows Alice to modify her brain. In fear of making irreversible damage to herself, she corresponds with Bob to develop a safety protocol, such that in the event Alice starts to feel pain, Bob can press a button to bring the old Alice back. Alice starts Neuromod, which unintentionally causes the following changes to her brain:

1. Neuromod makes Alice experience pain
2. Neuromod makes Alice not able to express the pain

Without the safety protocol, Alice would continue living with pain without being able to express this to Bob. In the cause Neuromod makes other advancements, it might happen that Bob decides to use Neuromod too, without Alice able to warn him. As a result, the two might end up with permanent undesirable experiences.

Luckily, the safety protocol works, Bob pushes the button, and Neuromod reverses the changes.

How does this safety protocol work? What kind of input does Bob receive to push the button?

While Alice and Bob's experiences of the world are profoundly mysterious to themselves, they agree to about the behavior of the system of experiences.

For example:

- when Alice perceive a red color, she writes down a number between 0 and 1
- 0 means no perception of red color
- 1 means the most extreme perception of red color Alice has experienced

Bob does the same for what he perceives as a red color, and they have now two scales to compare.

Having two scales to compare does not mean that Alice and Bob have a unique way of translating any experience among themselves, because e.g. 0.7 for Alice could mean 0.3 for Bob. The only fixed points so far are 0 and 1. Still, the closer a number is to 0 and 1, the more likely is that Alice and Bob means approximately the same thing. Therefore, when they both see a red color, they will agree they are seeing a red color, and expect to predict accurately whether the other is seeing a red color in the future. In addition, they perform tests to check how observing the same phenomena is evaluated at the two different scales. This gives them a way to interpolate the results such that they can predict experiences accurately in the other person that none of them has perceived before.

Alice and Bob divides and measures experiences into such scales, which they call “qualia”. The word qualia refers to the unknown sensation they both have, but the functional mapping between qualia is known. Therefore, functions of type `qualia → qualia` are known, but since they can not compute with qualia directly, they replace these functions with numbers:

experience[measurement] : real → real

experience : qualia → qualia
measurement : qualia → real

The laws for qualia that Alice and Bob agrees on are the following:

1. A functionally unmodified brain experiences similar qualia over time under repeated conditions
2. If there exists a brain experiencing A, and a brain experiencing B, then there exists a possible brain capable of experiencing A and B

With other words, unless Alice uses Neuromod, she will continue experiencing the world the same way as before. In principle it should be possible for her to use Neuromod to learn how to see a new color, while keeping her current set of experiences in mostly unmodified form.

This internal language of describing the nature of experiences follows the concept of intentional paths in path semantics.

An intentional path is a property of a function that can vary for logically equivalent cases.

Bear in mind that a function is per definition equivalent to another if it returns the same output on the same input. When talking about functions of the brain, it is custom to mix up the logically equivalent behavior with the properties of the specific embodiment of that behavior. Since this can get confusing, one can instead use path semantics to cleanly separate these two concepts.

The intentional path of `f` by `g` is written:

$f^g : A \to C$

$f : A \to B$
$g : (A, f(A)) \to C$

When Alice and Bob see a red color, they are processing information in a way that only uses a fraction of their total ability to see colors. One could say that there is an associated qualia with the specific processing of `(x, red)` of type `(object, color)`.

When Alice’s brain outputs “red”, it does not mean that Alice experiences the description “red”. It is more accurate to say that Alice can refer to her experience as “red”, because the brain outputs “red” to later recall that experience from memory. Using descriptions of experiences, Alice’s brain can reverse the process and associate objects together that causes similar effects of qualia.

A consistent way of generating qualia makes it easier for the brain to take advantage of memory. At any given moment, the only experiences Alice have consist of qualia, therefore any motivation to make predictions about the future is to achieve a desired stable pattern of qualia. Logical equivalence is necessary for memory, planning and communication, but the intentional equivalence is what gives Alice meaning to her existence. It is the intended aspect of her brain functions.

$(x, \text{red}) \rightarrow \text{qualia for “red”}$

Alice might have different moods which allows her to perceive a red color in different ways. This corresponds to different mappings of `(x, red)` into qualia. Notice that only the function input is needed when assuming the output is determined by a logically equivalent function.

$\text{vision}^{\text{alice0}} : \text{object} \rightarrow \text{qualia}$
$\text{vision}^{\text{alice1}} : \text{object} \rightarrow \text{qualia}$

$\text{vision} : \text{object} \rightarrow \text{color}$
$\text{alice}_0 : (\text{object}, \text{color}) \rightarrow \text{qualia}$
$\text{alice}_1 : (\text{object}, \text{color}) \rightarrow \text{qualia}$

For both cases `$\text{alice}_0$` and `$\text{alice}_1$`, she assigns 1 to the most extreme perception of red, because otherwise these two embodiments of the function `vision` would not be logically equivalent to each other. Still, it does not mean that these two perceptions of red are the same. The qualia information is not an isolated and independent immaterial state, but available to Alice’s brain through self-reflection, such that another part of her brain can perceive the difference between the two ways of perceiving red. Starting with two logically equivalent functions `$\text{vision}^{\text{alice0}}$` and `$\text{vision}^{\text{alice1}}$`, under consistent self-reflection these functions can be modified into two non-equivalent functions `$\text{vision}_0$` and `$\text{vision}_1$`.

$\text{vision}_0 : \text{object} \rightarrow (\text{color}, \text{real})$
$\text{vision}_1 : \text{object} \rightarrow (\text{color}, \text{real})$

Alice’s brain has now reduced an internal experience into a computation with an internal consistent language relating the two kinds of experiences to each other. This is kind of the same thing that happens when Alice and Bob measure and compare their experiences. They create an external system that is capable of making predictions about the natural way for brains to process information.

For example, for `$\text{alice}_0$`, a 1 could mean a 0.7 for `$\text{alice}_1$`. Since 1 for `$\text{alice}_1$` is a more extreme perception of red than what is possible for `$\text{alice}_0$`, one could create a function that normalizes over the maximum value and then use the new scale to describe both. This allows a way of reasoning about the combined experiences, as if you are making predictions about a brain that could perceive both.

This follows from the rule “If there exists a brain experiencing A, and a brain experiencing B, then there exists a possible brain capable of experiencing A and B”. The rule does not tell how to construct that brain, but this problem can be solved by reverse-engineering the brain.

Using Neuromod as a device to change her brain, one can think of the safety protocol to prevent damages as the following:

1. There is a database of how Alice measures experiences
2. There is a list of rules that describes how experiences are related
3. There is a list of desirable and undesirable experiences

If the change made by Neuromod predicts an undesirable experience, Bob pushes the button to restore the old Alice. The system that Alice and Bob developed does not solve the problem “what substances are qualia made of”, but “how to make predictions about experiences” such that Alice does not have to inform Bob about the result of a change to her brain. The system predicts Alice’s experiences as if Alice was still able to report them.