

## Consciousness and the Relativity of Science

# or Solving the Hard Problem of Consciousness (in 22 Easy Steps)

**1** Suppose we look inside the human cognitive system and notice the following. The system has a mechanism (call it the **analysis mechanism**) that is responsible for handling questions like “What exactly is the nature of [x]?” ... but when [x] is a color quale like [red], the analysis mechanism gets into trouble, because the [red] concept is directly attached to an incoming signal line and has no precursors. The [red] concept cannot be unpacked the way a normal concept can.

**2** However, the analysis mechanism does not report a failure, it simply returns an ‘answer-concept’ that points at nothing more than [red] did. As a result, the cognitive system eventually reports that “There is definitely *something that it is like* to be experiencing the subjective essence of red, but that thing is ineffable and inexplicable.”

**3** This “failure” of the analysis mechanism is common to all the consciousness questions. For qualia, the mechanism dead-ends into the sensory system. For the concept of Self, there is an innate representation for the self that cannot be analyzed further. All subjective phenomenology is associated with such irreducible concepts. Moreover, this is a **structural failure**: any cognitive system worth its salt would have the same problem.

**4** Now suppose that, on further investigation, we discover that all philosophical investigations that have ever been built on these questions turn out to be dependent on the same low-level failure of the analysis mechanism – all arguments eventually use this same appeal to intuition at some point, and as a result all arguments depend on this low-level, raw failure of a particular mechanism in the cognitive system.

**5** This is an account of why all consciousness questions are answered the way they are, and why we would expect there to be some strange-looking concepts in our minds that purport to describe a real thing (like the essence of red), but which look suspiciously like fictitious or broken concepts. To some, this might count as an explanation of consciousness.

**6** Now an important question for you, the person reading this poster. Does this argument sound convincing as an explanation for consciousness? Or do you feel that this is only an account of the psychological mechanisms involved in the answering of questions about the nature of concepts, and that something has been left out?

**7** Assuming you do feel this explanation is not sufficient, here is a further question. If you knew that the conceptual structures you built in your mind when you were reading the above argument had just been fed into a set of “evaluation” mechanisms, and if you knew that these mechanisms gave a thumbs-down *for exactly the same reasons that the analysis mechanism itself failed*, would this affect your conclusion?

**8** This approach can be developed to make some testable predictions (see below), and in the future it might lead to a series of arbitrarily fine-grained accounts of consciousness. This is reminiscent of the **Dirac delta function**: a pathological non-function that cannot be used by itself, but which can be used if it treated as the limit of a series of normal functions.

**9** Prediction: **New Qualia**. Make some new color receptors in the eyes, which are sensitive to IR. They should give rise to a new color quale. Then swap connections on the red and IR receptors, then remove the IR receptors and (old red) pathways: the old red quale will disappear.

**10** Prediction: **Synaesthetic Qualia**. Take the system from 9 and arrange for a cello timbre to excite the old lines that would have excited red qualia: cello sounds will now cause the system to have a disembodied feeling of redness.

**11** Prediction: **Mind Melds**. Join two minds so that B had access to the sensorium and concepts of A, using new pathways in B’s head. B would say that she knew what A’s qualia were like. If you use B’s *existing* sensory pathways, however, B would say A’s qualia were the same as hers.

**12** Prediction: **Animal Minds**. To experience what it is like to be an animal, you would need enough extra brain to hold a copy of the animal mind alongside your own. The experience would only be ‘genuine’ if you switched off your mind. Integrating the minds afterwards would be difficult.

**13** Naive conclusion from this analysis: **consciousness might, after all, supervene on the physical**. Surely there is no logically possible world in which the facts about the human cognitive system could be the same, but the experience of consciousness could be absent?

**14** Objection: we did not refer to the supervenience characteristics of the *experience* at all, only some of the psychological mechanisms associated with the experience. We slipped from Hard Problem to Soft Problem.

**15** True, but notice that we claim the possibility (at least) of a series of explanations that will eventually box the consciousness problem into a corner, and the thing in that corner might look very much like a failure in the concept system.

**16** Objection 2. The argument hinged on an *a posteriori* concept of consciousness, evading the primary intension of conscious experience and skipping forward to a better developed, secondary intension. And only primary intensions count in supervenience claims.

**17** This is new knowledge, not a secondary intension of consciousness. You can **unknow** that water = H<sub>2</sub>O without damaging the primary intension of water, but the facts about consciousness introduced here are too important to unknow without going into denial about something crucial.

**18** **The Solution (1)** What is the nature of “explanation” anyway? In no other case of explaining concepts (in physics, etc.) does **the implementation of concepts in the mind of the explainer** ever become part of the explanation. The old rules of explanation simply do not apply.

**19** **The Solution (2)** In science and in thought, what validates concepts is a **consistency game** that they play inside minds. As a result, concepts are separated into a class of REAL concepts and a class of NON-REAL concepts, according to how well they play this game.

**20** **The Solution (3)** Consciousness concepts, although apparently real, do not even take part in the game of consistency: they are a unique third class of concepts, neither real nor non-real but strictly outside the bounds of science. They are relative to nothing, and so exempt.

**21** Science contains, at its heart, a ‘dead zone’ where it can never tread. It can predict the existence of consciousness, and talk about some of its characteristics, but never actually say anything about its essence. This is both an explanation and not an explanation, depending on how you see it: it is not an explanation because it avoids the core, but is also a strange kind of meta-explanation because it explains why it cannot explain the core.

**22** A computer designed in such a way that it had the same problems as we do with the analysis mechanism (arguably, any computer intelligent enough) would experience consciousness. In the case of the computer, its mental life would be just as mysterious as our own, so our ability to build an intelligent computer would not condemn human minds as “mere” computers.

Richard Loosemore  
(Singularity Research Institute)  
Genoa, NY 13071  
rpwl@lightlink.com