Loosemore, R.P.W. (2006). Consciousness and the Relativity of Science.
Poster presented at the 2006 Toward a Science of Consciousness Conference, Tucson, AZ.

Explaining Consciousness by Explaining That You *Cannot* Explain it, Because Your Explanation Mechanism is Getting Zapped

or

Consciousness and the Relativity of Scientific Explanation

Richard P. W. Loosemore

~~RPWL@lightlink.com~~
rloosemore@susaro.com

> "I checked it very thoroughly," said the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is."
>
> *The Hitchhiker's Guide to the Galaxy*
> Douglas Adams

1 Suppose that one day you look inside a philosopher's cognitive system when she is asking herself "What is the essential subjective experience of redness?", and you discover that her ***analysis mechanism***, which would normally be responsible for unpacking a concept like [red], to track down its precursors, is hitting a snag. Normally the analysis mechanism would find the other concepts that the [red] concept points to, but because [red] is directly connected to a sensory input channel, the analysis mechanism runs into a dead end.

2 Looking further, you notice that the analysis mechanism responds to the crisis in a dumb way: it pretends that nothing is wrong, and returns a concept that represents [the meaning of red] – but where this concept should have had pointers to its constituents, there is nothing. It is a barren concept, a strange exception in a system where all concepts are supposed to point to the things that constitute them.

3 As a result of getting back this peculiar concept, the philosopher finds herself saying (a) there is definitely something that is the essence of redness, but (b) this thing is ineffable and inexplicable. This peculiar concept is certainly valid (it behaves just like all the others in the system), but it is empty.

4 Strangely enough, it turns out that *all* consciousness questions involve concepts of this sort. When the analysis mechanism tries to unpack any of them, it hits a dead end. (Think about it: all the hard problems involve primitives that have no precursors).

5 In the normal run of things inside a human cognitive system, there are some routine mechanisms that give the stamp of approval to concepts that are deemed "real" or "valid," or which have *bona fide* "thinghood," and when a concept has *strong* connections to the rest of the network of concepts, it is deemed *more* real. The peculiar [the meaning of red] concept gets high marks for reality because it is attached to [red], which gets 100% marks because of its primitive link to a sensory channel. So whatever happens in our cognitive system that ordinarily gives us a gut-level feeling for what is "real," that same mechanism classifies [the meaning of red] as real.

6 Any cognitive system built along the above lines would say, feel and think exactly the same things that we do about consciousness concepts. It would build philosophical arguments that were all predicated

upon this one failure of the analysis mechanism, and those arguments would purport to show that consciousness was a real thing in the world that could not be explained or reduced.

7 Does the above argument feel, to you, like it only discusses functional aspects of consciousness without addressing the hard problem? And if you say "yes", are you BEING an analysis mechanism when you give that answer, or are you thinking about being an analysis mechanism...?

8 Inasmuch as the [the meaning of red] concept is real, there is nothing we can say about what it is. But what we can do is understand *why* we can say nothing. We cannot explain consciousness, but we can explain why we cannot explain it.

9 Nowhere else in science does the content of an explanation involve aspects of the actual thinking mechanism that is trying to do the explaining. Whatever formalism we might try build to capture the meaning of "reductive explanation" in the context of explaining physical stuff out there in the world, that formalism cannot be simply imported into a situation where the subject of study involves a (possible) malfunction of the explanation mechanism. In particular, the notion of *supervenience* falls apart.

10 Ultimately, a case could be made that "Explanation" is not an objective thing (like collections of functions on possible worlds) – it is nothing more nor less than ***what explaining systems do***. Explaining systems like the human mind are quite possibly complex systems, so any regularities in the functioning of the explanation system might not be capturable as a formal system. Science, from this point of view, is not about uncovering the hidden nature of the universe, it is about an elaborate and limited construction built by human minds.

11 This argument does not claim that consciousness is just an artifact. All our "real" concepts are grounded on the reality of these primitive concepts, so classifying the latter as "not real" would infect all the others. In that case, science needs three classes of concept, not two:

(Type A) real & consistent (e.g. [proton])
(Type B) nonreal & inconsistent (e.g. [phlogiston])
(Type C) quasi-real & inexplicable (consciousness)

Type C concepts are just as "real" as any others, but they are strictly beyond the reach of science.

12 In spite of this, we can make some testable predictions about many aspects of phenomenal consciousness:

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.

Synaesthetic Qualia. Take the system above 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.

Mind Melds. Join two minds so that B has access to the sensorium and concepts of A, using new pathways in B's head. B will now say that she knows what A's qualia are like. If you use B's existing sensory pathways, however, B will say A's qualia are the same as hers. This is the only way to compare subjective experiences: the result depends on how you choose to do the cross-wiring. So the Inverted Qualia question is moot: the comparison of one mind's qualia with another's is strictly meaningless unless you actively cross-wire them.

Conclusion

Both sides were right. The core experience of conciousness cannot be explained. But nevertheless we need nothing but physical science and a functional model of mind to explain the existence of consciousness, and to explain why we cannot explain the core.