

International Journal of Machine Consciousness  
 © World Scientific Publishing Company

## What Does A Sloman Want?

DREW McDERMOTT

*Computer Science Department, Yale University, P.O. Box 208285,  
 New Haven, CT 06511-8285, USA*  
 drew.mcdermott@yale.edu

As often happens, I want very much to agree with Aaron Sloman. I share his opinion that there is not much productive research to be done on constructing conscious machines at this stage in cognitive science. I also believe (and who doesn't?) that we are mostly confused about the concept of consciousness; and that the word covers a multitude of different capacities whose differences will in the future be much more clearly understood, but only after a lot of preliminary work.

If the paper had stopped after section 1, the paragraph above would have been my happy summary of a position that amounts to "Hear, hear!" Unfortunately, the paper goes on, and repeatedly defies its earlier crisp deprecations.

Philosophers who stray from their roots often seem so relieved to be free of the necessity of weighing every word carefully, splitting hairs that are already submicroscopic, and finding arguments with numbered premises for every pronouncement, that they veer to the opposite extreme and say pretty much anything that pops into their head. Thus it can be quite difficult to grasp what a Dennett, a Rorty, or, alas, a Sloman really means. One can find in their writings quotes for and against many stances we would dearly love to know whether they want us to take or shun.

Section 2 proposes that once we have accounted for all the "competences" supposedly involving some aspect of consciousness, then we will understand all there is to understand: "...[W]e shall have solved the only problems of consciousness worth solving." This would have been another good place for the paper to stop.

But the same thought is repeated at least three more times. Once we have abandoned the label "consciousness," and "explain[ed]" or "specif[ied]" "every *other* fact about human mentality" (sect. 2), or "human minds" (sect. 9), then

- (1) "there will be nothing left to be explained about human consciousness" (sect. 2)
- (2) "it is not clear what will be left out that could be added that would make any difference to anything" (sect. 9)
- (3) "all substantive questions about consciousness...will have been answered, whether philosophers agree or not" (later in sect. 9)

But by this time it is quite unclear what we're being asked to agree with, since the list of things to be explained includes perception, "many affective abilities," "the ability to have... desires, inclinations, and preferences," and "readiness for [hypothetical] contingencies" (sect. 9). Furthermore, even the seemingly clear idea of a "competence" includes the ability to "think about numbers" (sect. 2), which many people, including me, can get quite muddled about if "thinking" is understood as nonconscious.

It becomes clear that we have a problem: if we don't understand or agree on what consciousness is, how can we be sure that we're *not* working on it?

In sect. 7 of the paper, all coherence is lost when the paper proposes that, "For a philosophical robot to start thinking about the phenomenal contents of experience, or 'qualia,' its introspective mechanisms would need to be able to access and record the contents of at least some of the perceptual processing subsystems." Isn't this exactly the kind of claim we are supposed to refrain from making for the indefinite future?<sup>a</sup>

I also don't get who, besides "philosophers," is being inveighed against. (Surely not the *International Journal of Machine Consciousness*!) Very few people in the AI community are working on consciousness, under that label or any other. Most are laboring away the way they always have, looking for opportunities to generalize programs that tackle intractable problems. (One such opportunity is the incredible mine of information available on the Internet. This is where I expect progress to be made in the near future.)

In sect. 3, Sloman quite sensibly says, "What needs to be explained is rarely evident at the start of a scientific investigation..." We (the cognitive-science community) will be blundering around for quite a while before we have understood what the relevant concepts are, and none of them may answer to our current terms, such as "consciousness." But doesn't the last paragraph of section 4 contradict this methodological point?

In order to make progress [on designing intelligent systems], we require, but currently lack, a good set of concepts for describing and comparing different sets of requirements and different designs: we need ontologies for requirements and designs and for describing relations between requirements and designs. . . . Without such a conceptual framework we cannot expect to cope with the complex variety of biological designs and the even larger. . . space of. . . artificial designs. Unfortunately, . . . different terms are used by different researchers to describe architectures, capabilities, and mechanisms, and often the same word is used with different interpretations.

In my view, and what I thought was the view of sect. 3, trying to correct the conceptual and terminological confusion *in advance of trying to design intelligent*

<sup>a</sup>A very charitable reading would be that this idea of a "philosophical robot" is a migration from sect. 12.

*agents* would be a complete waste of time.

The proposed Turing-style test in sect. 12 is clever, but I really hope it's not meant seriously. For one thing, it perpetuates the tradition of making up *a priori* behavioral tests that at best provide sufficient conditions for the proposition to be tested, a tradition that has been a huge distraction for the cognitive-science community since Turing, probably inadvertently, started it. The very idea that a scientific field based on opening up the head and modeling what's inside would need a blackbox test for whether an entity is intelligent is ridiculous and, after sixty years, tiresome.

The new test is also yet another way for Sloman to have his cake and eat it, too. We're not supposed to worry about consciousness; but we *are* allowed to make up tests that *explain* worries about consciousness. Isn't this just a sneaky second-order approach to studying the prohibited subject?

To summarize: I agree with everything the Puritan Sloman says in this paper. If only we could convince the Royalist Sloman to stay away.