Commentary on
   Max Velmans, 'Is Human Information Processing Conscious', in
   Behavioural and Brain Sciences C.U.P., 1991

Developing concepts of consciousness
Aaron Sloman
School of Computer Science, University of Birmingham

Velmans cites experiments undermining hypotheses about causal roles for consciousness in perception, learning, decision making, and so on. I'll leave it to experts to challenge the data, as I want to concentrate on removing the surprising sting in the tail of the argument.

Having argued that consciousness has no functional role, he does not conclude that it is just a myth (like the aether). Instead, he argues that because there is no functional "third-person" status, it must be given some other kind of "first-person" analysis. He also says that 'Consciousness is, nevertheless, amenable to scientific investigation', and that a complete psychology requires two complementary, mutually irreducible, perspectives, one of which studies consciousness.

Curiously, despite discussing subjects' reports on conscious episodes, he does not consider the possibility that the functional role might include enabling individuals to give others information about their mental states (useful in many contexts: in the family, in the classroom, in the dentist's chair, etc.)

I won't defend this view because, as explained below, the word "consciousness" is associated with too many muddled ideas for any such statement to be worth defending. Instead, I'll propose a different approach, from an engineering standpoint (Dennett's 'design stance', Dennett 1978). This views humans (and other animals) as having a complex and sophisticated design (which is not to say that there's any designer), and attempts to consider how one might make something with similar capabilities. Design provision must be made for "consciousness", "awareness" etc. (See also Dennett 1983)

Unfortunately it is not at all clear what this means. Because "consciousness" is ill-defined, Velmans says '...it is consciousness in the sense of "awareness" that is of primary concern'. Is "awareness" any clearer? Is the fly aware of my approaching hand? Is a dreamer aware of the pain, or the pursuing lion? Most people say: "Yes, that's why dreams are nice, or frightening." Others might think you can't be aware (= conscious) when you are asleep (= unconscious). Is the sleepwalker aware of the door-handle when he looks at it and turns it? We are all aware that the end of the century is approaching. If this is included in 'the sense of "awareness"' then consciousness includes nearly everything we know. Where are the boundaries?

Velmans is apparently mainly concerned with *self* awareness, i.e. inwardly perceiving one's internal states and processes. (Perceiving them, not just knowing about them?) There are many situations in which we are self-conscious or self-aware in some way. But not throughout waking life: when you are totally absorbed in something outside you, e.g. watching an exciting football match, or a gripping play, do you then lack "first-person" consciousness?

Conjecture: people who discuss consciousness delude themselves in thinking that they know what they are talking about. I don't claim that there is *nothing* they are talking about. Rather, it is not just one thing, but many different things muddled together. That colloquial language uses one noun is no more evidence for a unique reference than the multifarious uses of the word "energy" (intellectual energy, music with energy, high energy explosion, etc.)

Why not, like physicists, ignore colloquial usage and agree on some technical definition of the word "consciousness"? Partly because there is also considerable emotional energy associated with the word, which will interfere with serious usage of any technical homonym. More important, we don't yet have an adequate understanding of the issues: We don't know what the relevant capabilities of human beings and other animals are; we don't know what functional decomposition (i.e. what sort of architecture) underlies these abilities, and we don't know what sorts of mechanisms (electrical, chemical, neuronal, software, or whatever) are capable of producing such functionality.

Claiming to know what consciousness is by attending to it is no more convincing than claiming to know what spatial locations are by attending to them. It didn't help Newton. Attending doesn't answer questions about identity: "When is another thing the same place (or mental state) as the one attended to?" has different answers depending on what relationships are in question. A fly, a mouse and a person may all be aware of a moving object: Is that the same state?

There's no answer because there's nothing unique that you have and others definitely do or do not have. I am not denying the existence of what's attended to -- just its unique identification. Your state is very complex and other things may have states that are partly similar, partly different. But in what ways? How many different substates underlie "conscious" states? What feels like something simple that is either present or absent, is actually something complex, aspects of which may be present or absent in different combinations.

If we give up the idea of a unique referent, we can instead survey relevant phenomena, analyse their relationships to other capabilities, and then attempt to come up with explanatory designs: a hard task. Psychologists, philosophers, linguists, anthropologists, hypnotists, artists, and others can explore, survey, and summarise the many and varied phenomena. Cognitive designers can work both bottom-up trying to extend existing computational models and top-down, trying to produce detailed requirements analyses and design specifications for systems with human-like capabilities. Then, after analysing the design-tradeoffs, we can try devising mechanisms capable of generating all these capabilities, including self-monitoring capabilities.

An architecture that supports not only the perception of external events but also the monitoring of relatively global and abstract *internal* states could have a number of features consistent with the data reported by Velmans.

If sensory mechanisms monitoring the environment can themselves be monitored, there will be a delay between the occurrence of the first-order perceptual processes and the results of second-order monitoring, just as there is a delay between retinal events and the production of information about the environment. Moreover, self-perception, like all perception, will involve errors, loss of information, distortion, and so on.

Similarly, some designs support the ability to monitor "control processes" that take decisions, form plans, initiate actions. Again one can expect delay between the occurrence of the first-order internal events and the production of high level summary information about those events, and perhaps also some distortions and errors.

It is not surprising that results of monitoring that occur after the events monitored cannot be causally involved in their production, but they can still have a causal role: informing others, keeping records of internal processes for many purposes, including long-term feedback that revises strategies used in first-order processes. Experiments showing that *some* kinds of learning occur without high level self-monitoring do not imply that *all* do.

Another second-order process is high level decision making. The various subsystems that produce or control actions of various sorts may themselves be subject to "meta-level" control. Many parts of the system will normally chug along on the basis of information available to them. But some mechanism is required for coping with conflicting needs and for high-level long-term coordination and strategy formation.

This might use a "democratic" voting system with numerical summation and comparison procedures (e.g. in neural nets). Alternatively, high-level strategy formation and conflict resolution might be reserved for a special subsystem with access to more information and more powerful reasoning capabilities than the others (Sloman 1978). Mixed modes of global control are also possible.

Training could also use second-order processes. Many skills require low-level mechanisms to be trained by being taken through various steps in a complex process, with fine-tuning based on feedback (e.g. learning to drive a car, play the violin, pronounce words). This might use a mechanism that does partial analyses of the steps required, then guides the lower levels through those steps, and increasingly lets them take control. Internal monitoring of the behaviour of second (or higher) order control facilities would require yet another mechanism, alongside those monitoring perceptual processing.

These are but hints at the many and varied ways different levels of monitoring and control may coexist in intelligent agents: some mechanisms controlling others, some monitoring others, some training others, some resolving conflicts between others. Perhaps this somehow produces the illusion that there is *one* high level process in charge of and monitoring everything.

Conjecture: This (very difficult) design-based strategy for explaining phenomena that would support talk of consciousness will eventually explain it all. We shall have evidence of success if intelligent machines of the future reject our explanations of how they work, saying it leaves out something terribly important, something that can only be described from the first-machine point of view.

References
D.C. Dennett, *Brainstorms* Bradford Books and Harvester Press, 1978.
D.C. Dennett, 'Intentional systems in Cognitive Ethology', *Behavioural and Brain Sciences,* 6(3), 1983.
A. Sloman *The Computer Revolution in Philosophy: Philosophy Science and Models of Mind,* Harvester Press, and Humanities Press, 1978.