Available online at www.sciencedirect.com
SCIENCE DIRECT®
Consciousness and Cognition 14 (2005) 115–118
Consciousness and Cognition
www.elsevier.com/locate/concog

Evolutionary pressures and a stable world for animals and robots: A commentary on Merker☆

Stan Franklin*

Computer Science Division, The Institute for Intelligent Systems, The University of Memphis, Memphis, TN 38152, United States

Available online 5 November 2004

Abstract

In his article on The Liabilities of Mobility, Merker (this issue) asserts that "Consciousness presents us with a stable arena for our actions—the world ..." and argues for this property as providing evolutionary pressure for the evolution of consciousness. In this commentary, I will explore the implications of Merker's ideas for consciousness in artificial agents as well as animals, and also meet some possible objections to his evolutionary pressure claim.

1. Autonomous agents and consciousness

Merker's work called to my mind the question of the possibility of non-biological agent being subjectively consciousness. Can a robot or a software agent ever be phenomenally conscious? Can Merker's ideas help with this question?

An *autonomous agent* is a system situated within an environment, and part of that environment, that senses its environment and acts on it, over time, in pursuit of its own agenda, in such a way

☆ Commentary on B. Merker (2003). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition*, 14, 88–113. This article is part of a special issue of this journal on the Neurobiology of Animal Consciousness.

* Fax: +1 901 678 5129.
E-mail address: franklin@memphis.edu.

1053-8100/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2004.09.003

that its actions may affect its future sensing (Franklin & Graesser, 1997). In addition to biological agents, including almost all animals, autonomous agents may be artificial. Among these are computer viruses, software agents and robots. Here we'll be primarily concerned with the latter two.

Any autonomous agent must continually answer the question "What do I do next?" (See Chapter 16 of my *Artificial Minds* on the *action selection paradigm* (Franklin, 1995).) Thus any animal that moves must continually select actions appropriate to its current environment. We conjecture that we humans complete a *cognitive cycle* to select such an action five to ten times during every second (Baars & Franklin, 2003)!

For an action to be appropriate, that is, life sustaining, its choice must be based on an accurate assessment of the current state of the environment. That is what *senses* are for. They support such assessments.

Now, consider any autonomous agent, be it animal or robot, that moves in space and has spatially sensitive sense organs attached to its body. By *spatially sensitive*, I mean that movement of the sense organ produces apparent movement at the surface of the sense organ, independently of any change in the animal's environment. Using one of Merker's examples, when I move my eye, the image on its retina changes rapidly regardless of what is happening in the environment. To make accurate assessments of the environment in the service of action selection, any such agent must be able to distinguish actual movements in its environment from apparent movements produced by it own movement of its sense organs.

Every autonomous, mobile robot, such as the Sony Aibo, will surely require spatially sensitive sense organs for moving appropriately in the world. Suppose, we build in mechanisms to shield the robot's action selection from apparent motion self-produced by its own movement of its sense organs. Such shielding mechanisms might conceivably be based on any of several different principles. One such principle would be to have such mechanisms allow the robot to construct its own individual, unified, coherent and stable world, as Merker argues that consciousness does for some animals. Might such a robot be subjectively conscious? It would seem at least possible. Can such mechanisms be designed? That is an empirical question for robot designers. Note how Merker's work gives direction to robot designers attempting to produce conscious robots.

Let us turn to the other principal type of artificial autonomous agent, the software agent. Such software agents typically "live" on operating systems within computer systems. One such system, IDA (Intelligent Distribution Agent) is an autonomous software agent developed for personnel work for the US Navy (Franklin, Kelemen, & McCauley, 1998) and used for cognitive modeling (Franklin, 2001). IDA's single sense modality recognizes ascii characters arising from email messages, operator system messages, or data base queries. This single sense supports IDA's action selection, which allows her to negotiate with sailors about new jobs in unstructured English. IDA can also deliberate and make volitional decisions (Franklin, 2000).

The IDA architecture implements Baars' Global Workspace Theory of consciousness (1988), and IDA is *functionally* conscious in that sense. Is she subjectively conscious? I have never thought so, but, until now, had no convincing argument against such a possibility (Franklin, 2003).[1] Merker's work allows one to argue that IDA is not subjectively consciousness because she does not move her single sense organ through space. The problem of apparent sensory change due to

[1] Philosopher David Chalmers (Chalmers, 1996) once suggested that I seek such arguments by claiming subjective consciousness for IDA, and challenging others to prove that she was not (personal communication).

bodily motion does not exist for her, though the necessity for continual action selection based on sensory data certainly does. (The idea of the cognitive cycle mentioned above originated from developing IDA.) Thus, IDA would have no need for subjective consciousness to correct for apparent motion. But, might she need it for other reasons.

This brings us to the issue of the functions of consciousness. Baars' *A Cognitive Theory of Consciousness*, Table 10.1 lists nine different tasks for consciousness (1988). Merker's task of shielding from the effects of apparent motion is not among them. However, you will find, implicitly, several other obviously critical tasks. Let me list a few of them in my own words.

1. Consciousness allows us to interact appropriately with our world via consciously mediated actions. I can navigate around the room without bumping into furniture. I can find the refrigerator door handle in order to open it. Many of the actions involved in these activities require conscious mediation.
2. Consciousness allows us to watch for unpredictable dangers. I need consciousness to see a truck bearing down on me as I step into the street.
3. Consciousness enables us to take advantage of unpredictable opportunities. While walking to a restaurant for lunch, I notice a mailbox and take the opportunity to mail the letter in my pocket.

Would one need subjective consciousness for such tasks, or would functional consciousness do? I think functional consciousness would suffice. However, subjective consciousness is one implementation of functional consciousness. That is, having subjective consciousness build for us a continuing unified, stable, coherent world as an arena for our actions enables the three tasks listed above. (IDA seems to provide an example of functional, but not subjective, consciousness.) Hence the three tasks listed above might also provide evolutionary pressure for subjective consciousness. Merker's evolutionary hypothesis is not the only one, nor does he claim it to be.

Several important philosophical questions arise concerning Merker's claims, and should be addressed. We conclude by posing and suggesting answers for them.

Would not it be possible, at least in principle, for an animal or a robot to produce a stable world-model without resorting to subjective consciousness? It would certainly seem possible in principle to me.

Would not it be possible to produce such a stable world-model using computational mechanisms without subjective consciousness? Again, this seems possible in principle to me.

Selection for a stable world-model seems reasonable, but why selection for subjective consciousness? My introspection tells me that my subjective consciousness provides just such a world-model. (I suspect the reader's introspection tells him or her the same.) My consciousness evolved somehow. Since it provides such a stable world-model, that would seem to be one likely evolutionary pressure. The other tasks of consciousness discussed above might well provide additional pressure.

Even if consciousness comes biologically with this world-model, it is hard to see why it is not just a byproduct, given that we can explain the behavior purely in terms of the computational mechanisms. If subjective consciousness is just a byproduct, we must address what produces such a byproduct. Some mechanism must be responsible, if it is in fact different from the mechanism

that produces the stable world model. Such a separate mechanism must have a cost, computationally or biologically. This seems unlikely to have evolved unless subjective consciousness has a fitness value. Gould and Lewontin argue successfully that "... such competing themes as random fixation of alleles, production of non-adaptive structures by developmental correlation with selected features ..." etc. might account for traits observed to have evolved (Gould & Lewontin, 1979). Can subjective consciousness be an evolutionary spandrel? I do not think so. I suspect that it is too central a trait, and likely much too costly in brainpower, to have evolved and survived without serving a vital purpose. Either an argument, or evidence, for subjective consciousness being a byproduct would be needed. I know of neither.

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Science, 7*, 166–172.

Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.

Franklin, S. (1995). *Artificial minds*. Cambridge MA: MIT Press.

Franklin, S. (2000). Deliberation and voluntary action in 'conscious' software agents. *Neural Network World, 10*, 505–521.

Franklin, S. (2001). Conscious software: A computational view of mind. In V. Loia & S. Sessa (Eds.), *Soft Computing Agents: New Trends for Designing Autonomous Systems*. Berlin: Springer (Physica-Verlag).

Franklin, S. (2003). IDA: A conscious artifact?. *Journal of Consciousness Studies, 10*, 47–66.

Franklin, S., & Graesser, A. C. (1997). Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent Agents III*. Berlin: Springer Verlag.

Franklin, S., Kelemen, A., & McCauley, L. (1998). IDA: A cognitive agent architecture. In *IEEE Conference on Systems, Man and Cybernetics*. IEEE Press.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London*, *Series B* 205, pp. 581–598.

Merker, B. (this issue). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition*.