How to make a Self

Pat Hayes
Institute for Human and Machine Cognition
University of West Florida http://www.coginst.uwf.edu/˜phayes

Abstract

The computational paradigm can account, in broad terms, for many of the phenomena of consciousness, as Dennett (1991) argues convincingly. The general idea is that the mechanism of consciousness is an internal representational narrative (also called a 'global workspace' or a 'world model'), suitably embedded in a functional architecture, and the phenomenal aspects of consciousness are the content of this narrative. However, this kind of account has some problems, most notably the fact that there seem to be aspects of this internal narrative which we are not, and cannot be, conscious of. The distinction between conscious awareness and unconscious information processing (for example in the visual system) must therefore be based on something more than the simple presence of the relevant information. Another missing aspect is the self; the sense of personal integrity which is a hallmark of normal conscious experience.

Building on pioneering work by Perlis (1997), we will suggest a representational solution to both of these problems. Perlis posits an "ur-quale" which arises from a particular kind of self-modelling computation. This paper sketches how strong self-reference might evolve in a simpler architecture as a result of the interaction of processes of truth maintenance, episodic memory and active spatial location, and why the resulting representational structure would give rise to many of the characteristic phenomenal aspects of the subjective self. This account has the merit of not requiring exotic computational architectures to support a self-concept. Finally we will suggest ways in which this mechanism might break down and produce psychoses.

References

D.C. Dennett, *Consciousness Explained*, Back Bay, 1991

D. Perlis, Consciousness as Self-function, in *J. Consciousness Studies* 4, 5-6, pp 509–25, 1997