# THE APPEARANCE OF CONSCIOUSNESS IN MACHINES

**Desmond Fearnley Sander, University of Tasmania**

[Abstract of a paper presented at the *Towards a Science of Consciousness* Conference, Skövde, Sweden. August 2001.]

Rational automata, implemented by the author and co-workers, engage in discourse with one another in correct English. Their conversation appears natural. They reason inductively and deductively, learn, remember and forget, and display human-like characteristics, such as curiosity, gullibility and duplicity. I surmise that rational automata will pass an objective Turing test — with the judge as passive observer, rather than participant. Rational automata, increasingly, will appear to be conscious.

Questions as to whether or not machines can be conscious (rather than merely appearing to be) can only be answered relative to an ontology in which consciousness is sufficiently well specified. The ontology of von Neumann and Wigner, elaborated by Stapp, promises to do this. It is broadly in harmony with how the world appears, and sees the presence of knowledge in the world as something that must be taken into account. For this theory to become compelling a clear account is needed of the nature of knowledge and its relation with information.

To paraphrase Stapp: "Knowledge is the currency of reality." I have substituted "knowledge" for his word "information", because it fits better in the quantum theory picture. I want to make a distinction between knowledge, as personal, immediate and incommunicable, and information, as impersonal, enduring and communicable. (In view of the baggage they carry, it might be better to invent new words.) The many reasons for making this distinction include that knowledge (unlike information) is positive and undeniable: what is experienced is simply there. Although one may utter the words "my experience is not real", the experience remains. It is with information that the classical symmetry between positive and negative arises.

Into each life comes knowledge of the world, immediate and fleeting. We may come to see consciousness as the process that each individual animate being engages in, of making its knowledge permanent, of feeding its unique experience into the shared world of information. In that case, classical computers would lack consciousness: they bring no knowledge into the world. What the implementation of communities of rational automata shows is that rationality does not require consciousness. Rather, rationality is a capability of performing computations, of processing information. The success, if we may call it that, of the human species springs largely from the fact that evolution has produced a form of life with this capability. Brains simulate computers.

The question remains whether this ontology is a good one: I say 'good' rather than 'correct', because ontologies, like religions, are to be distinguished by their efficacy. There are many empirical issues of physics and physiology that need to be decided before such speculations will compel belief, in the way that, say, Newton's theory of the motion of the planets compels belief. The purpose of my presentation is not to make converts. Rather, it is to suggest a form that a resolution of current controversies might take, and to urge scepticism toward evangelists: to leave room for the possibility that the appearance of consciousness in machines will not be what it seems.

*Topics:*
Ontology of consciousness. Machine consciousness.
*Keywords:*
Quantum; ontology; determinism; knowledge; information; computation; autonomy; language; inference; grammaticality; Turing test; free will; entanglement; decoherence.