

Artificial agents and the expanding ethical circle

Steve Torrance

Received: 7 December 2011 / Accepted: 31 March 2012 / Published online: 27 April 2012
© Springer-Verlag London Limited 2012

Abstract I discuss the realizability and the ethical ramifications of Machine Ethics, from a number of different perspectives: I label these the anthropocentric, infocentric, biocentric and ecocentric perspectives. Each of these approaches takes a characteristic view of the position of humanity relative to other aspects of the designed and the natural worlds—or relative to the possibilities of ‘extra-human’ extensions to the ethical community. In the course of the discussion, a number of key issues emerge concerning the relation between technology and ethics, and the nature of what it is to have moral status. Some radical challenges to certain technological presuppositions and ramifications of the infocentric approach will be discussed. Notwithstanding the obvious tensions between the infocentric perspective on one side and the biocentric and ecocentric perspectives on the other, we will see that there are also striking parallels in the way that each of these three approaches generates challenges to an anthropocentric ethical hegemony, and possible scope for some degree of convergence.

Keywords Machine Ethics · Artificial intelligence · Anthropocentrism · Biocentrism · Ecocentrism · Infocentrism · Moral status · Consciousness · The ‘more-than-human’ world

The following chapter was previously published, under the title ‘Machine Ethics and the Idea of a More-Than-Human Moral World’, in *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson, © 2011 Cambridge University Press, pages 115–137. It is reproduced here, with some amendments, and a new introductory section, by permission of Cambridge University Press.

S. Torrance (✉)
School of Engineering and Informatics, University of Sussex,
Falmer, Brighton BN1 9QJ, UK
e-mail: stevet@sussex.ac.uk

‘We are the species equivalent of that schizoid pair, Mr Hyde and Dr Jekyll; we have the capacity for disastrous destruction but also the potential to found a magnificent civilization. Hyde led us to use technology badly; we misused energy and overpopulated the earth, but we will not sustain civilization by abandoning technology. We have instead to use it wisely, as Dr Jekyll would do, with the health of the Earth, not the health of people, in mind’.
Lovelock 2006: 6–7.

1 Prefatory remarks: mental and moral universes

The following discussion emerges from a set of preoccupations that have coloured my thinking over three or more decades. My active interest in Cognitive Science and AI commenced in the early 1980s. I was fascinated, like many, by the philosophical implications that seemed to flow from the idea of computationally based intelligence, concerning the puzzle of what it was to have a mind. The idea that mentality was not necessarily confined to humans or other natural-born species was seductive and vertiginous. My own interest at that time in the potential expansion of the cognitive universe was strongly influenced by some reflections about ethics. If we are to expand the *mental* universe to admit machine-minds, I thought, does that not mean that we should also be expanding the *moral* universe?

My interest in the ethical implications of AI was connected indirectly with another field of ethics. In the mid-1970s I had come across Peter Singer’s *Animal Liberation* (Singer 1977), a book which provided a powerful, extended argument that humans have an obligation to take the suffering of non-human animals seriously from a moral point of view. The book seemed to offer a radical challenge to most received ethical and political viewpoints of the day. It