AI & Soc (2013) 28:483–489
DOI 10.1007/s00146-013-0453-3

25TH ANNIVERSARY VOLUME

A FAUSTIAN EXCHANGE: WHAT IS IT TO BE HUMAN IN THE ERA OF UBIQUITOUS TECHNOLOGY?

What is it like to encounter an autonomous artificial agent?

Karsten Weber

Received: 8 February 2013 / Accepted: 28 March 2013 / Published online: 20 March 2013

Abstract Following up on Thomas Nagel's paper "What is it like to be a bat?" and Alan Turing's essay "Computing machinery and intelligence," it shall be claimed that a successful interaction of human beings and autonomous artificial agents depends more on which characteristics human beings ascribe to the agent than on whether the agent really has those characteristics. It will be argued that Masahiro Mori's concept of the "uncanny valley" as well as evidence from several empirical studies supports that assertion. Finally, some tentative conclusions concerning moral implications of the arguments presented here shall be drawn.

Keywords Autonomous artificial agent · Turing test · Uncanny valley · Moral responsibility

1 Introduction

In his seminal paper "What is it like to be a bat?" published in 1974, Thomas Nagel argues in favor of the irreducibility of the first-person perspective. His main argument is that being a bat, a spider, an ape, or a human being implies that those particular living beings experience something that cannot be described or explained merely by using physical terms. Following Nagel, it is simply a fact that this or that living being has this or that experience; it is a fact that cannot be denied by arguing that such experiences can be described by merely using physical terms which do not provide for a space for an entity called mind or, in other words, private, respectively, subjective experiences named qualia. Roughly spoken, in his essay, Thomas Nagel decidedly rejects the idea of eliminative physicalism.

It would be possible to adopt Nagel's point of view to argue that if bats and spiders and all other living beings having something like a central nervous system or at least some large ganglia have subjective or private experiences—that "there is something that it is like to be that organism" as Nagel (1974, p. 436) puts it, then it would be a valid argument that there is something that it is like to be and for an autonomous artificial agent if this agent has some artificial equivalent to a central nervous system. If this assumption is true, then it also might make sense to ask whether such an agent actually is or might be morally responsible for its actions and the outcomes of its actions. However, even if this question would be possible to ask and would fit at least from a philosophical point of view in what follows, it shall not be discussed furthermore.

But this does not mean to simply abandon Nagel's essay and the intuitions he presents there. To the contrary, it shall be argued that Nagel not only rejects eliminative physicalism but also makes a claim that might be vital for further investigations with regard to autonomous artificial agents. Although Nagel's major aim is to argue in favor of the irreducibility of the first-person perspective, he also raises another issue that might be fruitful for the discussion concerning, whether autonomous artificial agents could be conceived as morally responsible actors. This argument or, better to say, this intuition is presented in merely one sentence in Nagel's essay. After some introductory paragraphs he writes the following:

> Even without the benefit of philosophical reflection, anyone who has spent some time in an enclosed space

K. Weber (✉)
Faculty 1, Mathematics, Natural Sciences and Computer Science, PO Box 101344, 03013 Cottbus, Germany
e-mail: Karsten.Weber@tu-cottbus.de