AI & Soc (2013) 28:431–441
DOI 10.1007/s00146-013-0500-0

25TH ANNIVERSARY VOLUME

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

Automatic decision-making and reliability in robotic systems: some implications in the case of robot weapons

Roberto Cordeschi

Received: 29 July 2013 / Accepted: 8 August 2013 / Published online: 28 August 2013

Abstract In this article, I shall examine some of the issues and questions involved in the technology of autonomous robots, a technology that has developed greatly and is advancing rapidly. I shall do so with reference to a particularly critical field: autonomous military robotic systems. In recent times, various issues concerning the ethical implications of these systems have been the object of increasing attention from roboticists, philosophers and legal experts. The purpose of this paper is not to deal with these issues, but to show how the autonomy of those robotic systems, by which I mean the full automation of their decision processes, raises difficulties and also paradoxes that are not easy to solve. This is especially so when considering the autonomy of those robotic systems in their decision processes alongside their reliability. Finally, I would like to show how difficult it is to respond to these difficulties and paradoxes by calling into play a strong formulation of the precautionary principle.

Keywords Autonomous robotics · Robot reliability · Robot weapons · Precautionary principle · Risk decision-making · Cybernetics · Internet

1 Introduction: when robots began to be regarded as autonomous

The notion of robot autonomy comes in varying degrees. Generally speaking—and for the purposes of this article at least—a robot is a (stationary or mobile) machine able to process data coming from its own sensors so as to interact with its environment to carry out a given task. At its most low degree of autonomy, a robot is a remotely operated machine, i.e. a machine remotely controlled by a human operator. Here, the robot can carry out some functions at least in an autonomous way, e.g. be guided by its own sensors towards a goal. At its highest degree of autonomy, a robot should be able more and more to interact successfully with unknown environments, also through its self-learning capacities, and in situations which its designer or operator had not planned for. Furthermore, such a robot should be able to take *decisions* on what to do and how to do it, without any "human in the loop"—i.e. fully autonomously from the human operator, or in a fully automatic mode. Robots like these are currently at the mainly experimental stage.[1]

When did we start thinking about "autonomous robots"? Some of the steps that, since the beginning of the past century, have led to the thought of robots as machines with differing degrees of autonomy from their designers or operators have been reported elsewhere (see Cordeschi 1991, 2002). This thinking came about in the 1920s and 1930s. The psychologist Clark L. Hull called the project to build machines with capacities comparable to living organisms such as adaptation and learning, hence

R. Cordeschi (✉)
Department of Philosophy, Sapienza University of Rome, V. Carlo Fea 2, 00186 Rome, Italy
e-mail: roberto.cordeschi@uniroma1.it

[1] Robot's decision or choice processes are usually considered the main hallmark of its "intelligence"—a term that refers to the ability of a machine to emulate cognitive abilities such as decision-making and learning, in the tradition of Artificial Intelligence (AI). In this article, I shall intend *automatism* and *autonomy* as strictly related terms, in accordance with some use of these terms (see in the following). In contrast with this, robotic *automatic* systems are seen sometimes as carrying out only fixed or preset operations (e.g. industrial robots, which are not "intelligent" in the aforementioned sense), and as such, they are opposed to robotic *autonomous* systems, endowed with the above-mentioned cognitive abilities.