Intelligence - Dynamics and Representations

Luc Steels

Artificial Intelligence Laboratory, Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
E-mail: `steels@arti.vub.ac.be`

Abstract. The paper explores a biologically inspired definition of intelligence. Intelligence is related to whether behavior of a system contributes to its self-maintenance. Behavior becomes more intelligent (or copes with more ecological pressures) when it is capable to create and use representations. The notion of representation should not be restricted to formal expressions with a truth-theoretic semantics. The dynamics at various levels of intelligent systems plays an essential role in forming representations. An example is given how behavioral diversity spontaneously emerges in a globally coupled network of agents.
Keywords. Intelligence, self-organisation, representation, complex dynamical systems

1 Introduction

Artificial intelligence research is concerned with an investigation into the phenomenon of intelligence using the methods of the artificial [24]. This means that systems are built which exhibit intelligent behavior and that this is seen as a way to progressively derive and test a theory of intelligence. After three decades of research, nobody denies that AI has resulted in many spinoffs for computer science, such as list processing, declarative programming, search algorithms, etc. Nobody denies that a whole range of programs have been written that exhibit features of (human) intelligence. For example, chess programs now compete at grandmaster level, expert systems have demonstrated human-level performance in difficult problems like scheduling, diagnosis, or design, natural language programs of high complexity have been built for parsing and producing natural language, and some machine learning programs have been capable to extract compact representations from examples. But substantial progress is still possible.

First of all, it seems that research efforts so far have not resulted in a coherent, widely accepted theory of intelligence. There is a body of engineering methods, techniques, and intuitive insights, which are usually taught using case studies [30]. This absence of theory is undoubtely due to the engineering bias of AI and the push from society to produce useful artefacts as opposed to theories. In addition, the bits of theory that do exist (for example Newell's cognitive architecture [19]) have no connection with theories of the physical and biological world, so that there is a wide gap between AI and other natural sciences.

L. Steels (ed.), *The Biology and Technology of Intelligent Autonomous Agents*

Second, there is a continuing criticism that the achievements of AI systems rest completely on the intelligence of the AI programmers: They extract and formalise the knowledge from experts, they set the conceptual framework and determine the set of good and bad examples for machine learning algorithms, they synthetise the grammars going into the natural language systems. The problem of how structures for knowledge and behavior may develop is largely unresolved. This is particularly a bottleneck in the area of sensory-motor intelligence and common sense, where the task of analysis, formalisation, and explicit programming is so formidable that little success has come from using the classical AI approach.

Current research in intelligent autonomous agents promises to tackle these two gaps in a fundamental way: It is seeking a theory of intelligence compatible with the basic laws of physics and biology and a theory which explains how intelligence may come from non-intelligent, material processes. Obviously we are far from achieving these goals. Only the contours of the theory are visible and at the moment the artefacts that can be built are promising with respect to sensory-motor intelligence but still weak with respect to 'higher level' cognitive tasks. But a new methodological track has been opened in which solid work can proceed.

The main pillars of the new approach are as follows:

+ A biologically oriented definition of intelligence is the starting point of the investigation. Intelligence is defined with respect to the capability of an autonomous system to maintain itself. This gives an objective criterion, as opposed to a subjective criterion based on judgement of performance or the ascription of knowledge and reasoning. This definition is refined by considering the functionalities used to increase the chances of survival: representation, specialisation, cooperation, communication, reflection, etc.

+ A theory of intelligence must be compatible with the basic laws of physics and biology and it must be a universal theory, i.e. independent of a particular embodiment (wetware or silicon) or system level (brain component, individual agent, society). This universality can be achieved by using complex dynamical systems theory as a foundation. Intelligence then is seen as the result of a set of non-linear processes which exhibit properties also found in other physical systems. Phenomena like behavioral coherence, cooperation, or the emergence of diversity between agents can be explained using bifurcation theory, chaos, self-organisation, etc.

The goal of this paper is to discuss these two directions of research in more detail. First we focus on the biologically oriented definition of intelligence (section 2), further refining it with the notion of representation (section 3). Then we discuss in which way dynamical systems theory can act as the theoretical foundation of a theory of intelligence.