Self-Aware Control Systems

Ricardo Sanz, Owen Holland, Aaron Sloman, Andrei Kirilyuk,
William Edmondson and Stephen Torrance

Background

Computer-based control systems in all fields (process control, avionics, robotics, etc.) are facing an enormous challenge. Trends in embedded software show a steady increase in size and complexity but the capability of the development technologies hasn't followed the needs and these days there is a manifest shortage of engineering capability to build the required systems of tomorrow. This is a very serious problem because complex information systems are not only required to provide increased levels of performance (for example in plant-wide optimisation) or enhance the constructability and maintainability of systems (for example in X-by-Wire systems) but also to increase or at least maintain the levels of dependability of the smaller embedded systems of the past.

Machines are becoming much more skilful thanks to the incorporation of massive doses of IT but deteriorating global system quality is a major threat that is limiting the application of the technology in some fields where dependability and/or adaptability is essential.

Valuable contributions follow from the proof-driven engineering approach but this has a serious flaw concerning our capability of designing and analysing of complex systems in the presence of state combinatorial explosion. Other alternatives that are being explored are those approaches based on the implementation of architectural mechanisms for self-organisation and self-repair. The system is constructed in a base state and it adapts based on the circumstances of the environment that surrounds the computing system. Sample exemplary efforts are those of the autonomic computing field but they lack the capability of incorporating reasoning about the physical world in the process of adaptation.

In some sense, the problems of the different approaches can be traced back to a deep deficiency of control systems: they do not understand the perceptual flow in the terms and up to the level necessary to provide the robust performance required for those systems mentioned above.

Research challenges

The objective of this research is to explore new architectural approaches in the design of intelligent controllers that attribute meaning to complex patterns of sensory stimuli and generate sequences of elementary actions that satisfy high-level goals. This should be done not only in terms of perception/action over the external world but also in terms of perception/action over the internal world, i.e. the body of the agent itself. An important problem is the proper characterisation of what constitutes the body of the agent and what constitutes the surrounding world (very different perspectives can be found today e.g. in mobile robotics and process control systems).

The ultimate aim is to build systems that exhibit flexible, autonomous, goal-directed behaviour in response to changes in internal and external conditions based on a deep understanding of the world and the self. They will have integrated control architectures that generate and exploit world- and/or self-awareness.

This research can achieve results of extreme importance in the construction of highly resilient, adaptable information systems that are urgently needed in many fields of computing and, in particular, in dependable autonomous systems. Some of the challenges for this research are:

- The analysis and design of the architectural foundations of integrated controllers with explicit representation and exploitation of “self”. This is expected to undo the fragmentation in recent science and technology of mind (put together mental pieces into “complete” architectures).
- Design of robust mechanisms for self perception and representation.
- Implementation of tools for architecture exploration by rapid prototyping and definition of test scenarios for evaluation of bioinspired self-aware systems
- Adequate and coherent interaction in the complex-dynamical system of environment, controlled plant, machine controllers, and human controllers.
- Mechanisms for active and passive screening of world- and self-perception.
- Performance analysis of “self”-based autonomous controllers.
- Design-centric vs. ontogenic approaches to self construction and exploitation. Constructive interaction between design-centric and ontogenic (conceptual, mathematical) approaches.
- Generation of theory and technology of synthetic phenomenology.
- Analysis of social, psychological, ethical and other philosophical aspects of the proposed engineering and scientific realisation of self-aware (conscious) machinery and the corresponding direction of development in the whole; necessary corrections and practical recommendations for the engineering solutions.

Research communities addressed

The research communities addressed by this focus point are those concerned with the construction of technical systems inspired by reflective and self-awareness control properties of biological brains and communities in cognitive neuroscience of consciousness. Some examples are:

- Autonomous Control Systems
- Autonomic Computing Systems
- Unified Theories of Consciousness
- Dependable Information Systems
- Dynamic Complexity in Embedded Systems

Ethical issues

This research is not focused on experimentation with humans and/or animals and hence it does not have ethical implications in this sense. The focus is on strictly technical matters and the only possible ethical issue is related to the remote possibility of achieving total success: i.e. being able to construct self-aware person-like synthetic systems. While this is not a concrete objective of the research proposed, it is obvious that if this research is totally successful the construction of synthetic persons should be considered as a possible application of the science and technology developed.

Powerful technological innovations tend to proliferate beyond control and if genuine artificially conscious systems became as widespread as PCs or mobile phones, questions of misattribution would be very serious. Also the consequences of a mass social misperception of artificial consciousness need to be considered. Such indirect social effects could be enormous, and difficult to predict. Like reproductive cloning, this involves ethical considerations that are quite novel, and difficult to focus on clearly. Ethical consequences of a novelty need not be negative, especially in this case, but any wide research programme should have a significant sub-project to investigate the ethical implications.