Intelligence, control and the artificial mind

Ricardo Sanz

Artificial intelligence and cognitive science must look at the world of industrial-process control to find the technological reifications of the concept of mind.

Artificial intelligence (AI) seems to be at an impasse. The old vision of AI which started as the search for a computer-based approximation of the human mind is not delivering. The initial hype opened the door to ample criticism following failures to fulfill some bold predictions. Cognitive-systems research (CSR) has replaced AI at the forefront of this research programme. But CSR is really just a new name for the same set of objectives, designed to elude the tag of failure. The problem with this programme may not be in the methods but in the naïve conceptualizations that have driven and are still driving the research.

Indeed, AI has not been a failure. Many AI technologies are routinely used with enormous success in domains from credit-card authentication to nozzle design and language understanding. And beyond the focused applications of concrete AI technologies, its big objective remains an ongoing success. However, the realization of AI is not to be found in the domain of robotics—still in its infancy—but in the uncontroversially materialistic and practical world of industrial-processing plants.

The challenges posed today by these complex technical systems set the proper stage for continuing the pursuit of the old dream of AI: the artificial mind. Current research delves into topics such as perception, understanding, self, and consciousness: not for human-like robots, but for plainly alien systems like refineries or electrical infrastructures. Intelligent control (IC) started as a process of technologically immersing AI into the world of control systems. For process control systems,[1,2] the availability of reusable inference engines led to implementation of expert systems exploiting the knowledge of human operators. At first, these systems were only usable as decision-support systems for humans. But with the development of real-time expert-system shells, one could use inference engines to implement closed-loop real-time controllers. At the same time, developments in fuzzy logic and fuzzy control technology enabled construction of systems embracing vagueness with better results than those obtained with other mechanisms
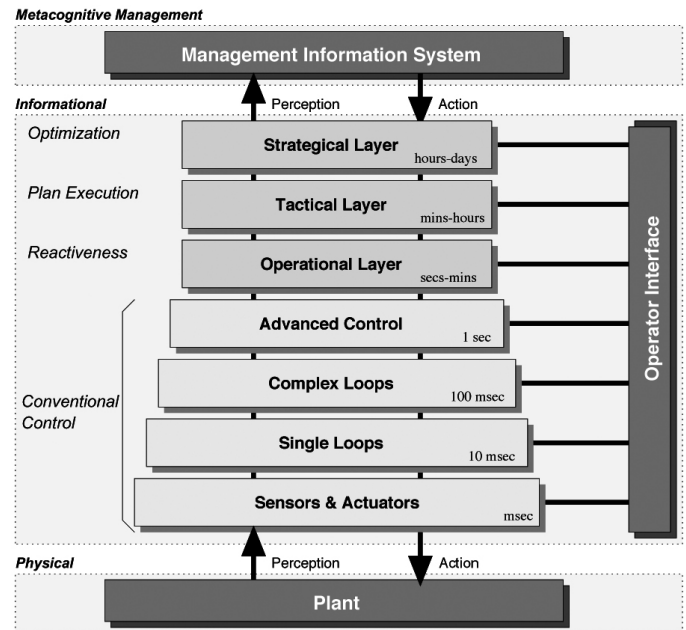

Figure 1. Typical functional layering in a complex industrial-process-control system.

such as bayesian or necessity-possibility frameworks. The same can be said about neural-network technology and its intrinsic learning capabilities. IC implies a systematic engineering path to the construction of automated operators, exploiting the knowledge of human operators and the deep plant knowledge of process engineers.

From simple, fuzzy rule-based systems at the lowest level to complex model-based reasoners at the strategic-control level, AI technology provides very effective mechanisms for improving controller competence in special circumstances (see Figure 1). The many claimed capabilities of the different AI methods provide major improvements on all scales of the control hierarchy, while the learning capability of nonlinear action mechanisms (neural networks, adaptive fuzzy controllers or genetic algorithms) was one of the key contributions of IC to the field of automatic control (AC).[3,4] However, the degree of predictability of the AI-based controllers was not as good as desired. This

obviously limited their use in safety-critical applications, but also raised justifiable criticisms when an expert system demonstrated brittleness or a genetic algorithm evolved truly stupid control rules. At the same time, the ad hoc approach used in most cases renders systems lacking the requisite property of robust autonomy.

IC quickly became a tool-driven endeavour instead of a problem-driven discipline. The research community gravitated around specific technologies which continue to be rule-based systems, artificial neural networks, fuzzy control and evolutionary programming, now classic subfields of the soft-computing world.[5,6] However, if we analyze the original motivations, we see that the control focus on AI was more than just a search for individual technologies.[7–9] It was a natural move because the control and AI communities were originally in search of the same objective: the technology of the artificial mind. In the case of AI, this was done as imitation of the human mind. For AC, methods of physics were used for any kind of body that was targeted. This common objective was much clearer in the past, such that AI and AC were just offspring of cybernetics.

Obviously, the many approaches of the AI panorama have not rendered the promised artificial mind.[10,11] But neither is the domain of AC so deeply trapped in the limited mathematics of linear systems. The clearest example is perhaps the humanoid robotics field. Where body dynamical control is achieving high levels of performance in bipedal walking, cognitive architecture is still very far from offering even a minimal glimpse of a real human mind.[12] The pursuit of the complete human-like mind was never an objective in the field of IC. It only sought some atomic human capabilities to improve performance of localized control systems.

The many successes of AI in control notwithstanding, at the very heart we still feel the lack of a technical capability to engineer the critical human competence of handling abnormal situations.[11] In both humanoid robots and intelligent controllers we feel the need of going beyond what we can do today and search for the seemingly missing essence of mind. This search has been a major pursuit in different fields—philosophy, neuroscience, psychology, robotics—which have converged into a single programmatic discipline: cognitive science. This is a heterogeneous community because of the many different backgrounds, research practices and personal research (sub)objectives. However, the emergence of a unified theory of mind is perceivable in the convergence of the theoretical models from the different domains. This unified vision is so powerful that it is providing a way to attempt to formalize
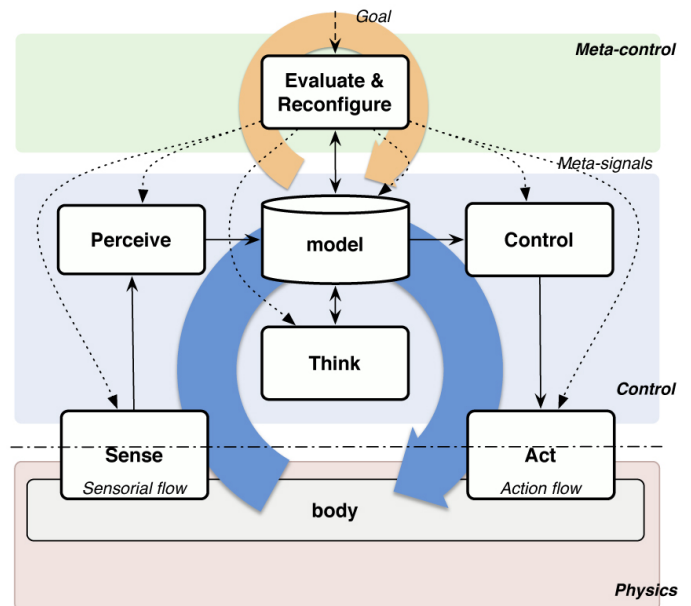

Figure 2. General atomic structure of the epistemic-control-loop pattern to be pervasively implemented across the control hierarchy.

such age-old conundrums as perception, knowledge, thought or even consciousness.

The IC community tried to mimic concrete human thought processes in its search for competence. The fragility of the realized systems calls for a new foundation which will not be found in the so-called 'new AI' or postmodern robotics. Cognitive science, on the other hand, is lost in the labyrinth of microdetails of the human mind and brain.

From our own research into these problems, we conclude that the only viable strategy to eliminate brittleness and increase mission-level resilience is to make systems epistemologically robust at the mission level,[13] so we can move the responsibility for real-time cognitive behaviour from us engineers to the systems themselves during runtime. And to do this we need what many think is the ultimate human trait: self-consciousness.[14,15] This is what we are trying to do with the development of the SOUL cognitive architecture for robust autonomy. One of the critical elements in this approach is the epistemic-control loop (see Figure 2), the basic design pattern for embedding intelligence pervasively into the system. A highly robust autonomous system will not only realize a hierarchical federation of cognitive-control loops but also a transversal metacognitive competence that will render the necessary self-awareness for

achieving full autonomy. We expect these conceptually rigorous artificial minds to be the theoretical cornerstones of a new science of mind. The artificial mind is coming.

Author Information

Ricardo Sanz
Autonomous Systems Laboratory
Polytechnical University of Madrid
Madrid, Spain

References

1. M. J. Stock, **AI in Process Control**, McGraw-Hill Inc., New York, NY, USA, 1988.
2. L. Boullart, A. Krijgsman, and R. A. Vingerhoeds (eds.), **Application of Artificial Intelligence in Process Control**, Pergamon Press, 1993.
3. R. Sanz and R. Galán, *A cooperative architecture for intelligent process control*, **Proc. Int'l Work. Conf. Coop. Knowl. Based Syst., CKBS 2000**, pp. 58–59, 1990.
4. R. Sanz, A. Jimenez, R. Galán, F. Matia, and E. A. Puente, *Intelligent process control: the CONEX architecture*, in S. G. Tzafestas ed., **Engineering Systems with Intelligence: Concepts, Tools and Applications**, pp. 407–413, Kluwer Acad. Publ., Norwell, MA, USA, 1991.
5. P. Bonissone, *Soft computing: the convergence of emerging reasoning technologies*, **J. Res. Soft Comp. 1** (1), pp. 6–18, 1997.
6. G. G. Yen, *Editor's remarks – a new page unfolded: IEEE Computational Intelligence Magazine*, **IEEE Comp. Intell. Mag. 1** (1), p. 2, 2006. doi:10.1109/MCI.2006.1597053
7. R. Sanz, I. Alarcón, M. J. Segarra, A. de Antonio, and J. A. Clavijo, *Progressive domain focalization in intelligent control systems*, **Cont. Eng. Pract. 7** (5), pp. 665–671, 1999.
8. R. Sanz, M. J. Segarra, A. de Antonio, F. Matía, A. Jiménez, and R. Galán, *Design patterns in intelligent control systems*, **Proc. Int'l Fed. Auto. Cont. (IFAC) 14th World Cong.**, 1999.
9. R. Sanz, *Agents for complex control systems*, in T. Samad and J. Weyrauch (eds.), **Automation, Control, and Complexity: New Developments and Directions**, ch. 10, pp. 171–190, John Wiley and Sons, 2000.
10. R. Sanz and I. López, *Minds, MIPS and structural feedback*, **Perform. Metrics Intell. Syst., PerMIS '2000**, 2000.
11. R. Sanz, F. Matía, and S. Galán, *Fridges, elephants and the meaning of autonomy and intelligence*, **IEEE Int'l Symp. Intell. Control, ISIC'2000**, 2000.
12. R. Sanz, F. Matía, and E. A. Puente, *The ICa approach to intelligent autonomous systems*, in S. Tzafestas ed., **Advances in Autonomous Intelligent Systems**, **Microproc.-Based Intell. Syst. Eng.**, ch. 4, pp. 71–92, Kluwer Acad. Publ., Dordrecht, Netherlands, 1999.
13. J. Gómez, R. Galán, R. Sanz, and I. López, *Naturalized epistemology for autonomous systems*, **Proc. KNEW 2007**, 2007. http://tierra.aslab.upm.es/documents/controlled/ASLAB-A-2007-014.pdf
14. R. Sanz, I. López, M. Rodríguez, and C. Hernández, *Principles for consciousness in integrated cognitive control*, **Neural Netw. 20** (9), pp. 938–946, 2007.
15. A. Meystel and R. Sanz, *Self-identity of complex control systems*, **Proc. Worksh. Theor. Fund. Intell. Syst.**, 2002.