Title:

Artificial Consciousness in Control Systems

Authors:

Ricardo Sanz

Address:

Autonomous Systems Laboratory
Universidad Politécnica de Madrid
José Gutiérrez Abascal 2
E-28006 Madrid
SPAIN

ricardo.sanz@aslab.org
Tel: +34 91 336 30 61
Fax: +34 91 336 30 10

Abstract:

Control systems are gaining in complexity to handle complex problems in uncertain conditions. Artificial Intelligence technology has served well to handle concrete problems in local scopes, but recent trends in controller construction lead to the need of plant-wide integrated systems in order to maximize efficiency. Surprisingly, Integrated Intelligent Controllers are reaching complexity level and behavioral features that touch that old big dream of the AI community: the conscious machine. Modern complex controllers are getting progressively conscious but not because it is fashionable, but because having a self is proving useful to achieve technical objectives.

Keywords:

Intelligent control, artificial consciousness

Topics:

- Fundamentos de la Inteligencia Artificial
- Control Reactivo, IA en Tiempo Real
- Robótica
- Modelos de Razonamiento

For the **Open Discussion Track**

Artificial Consciousness in Control Systems

Ricardo Sanz

Autonomous Systems Laboratory
Universidad Politécnica de Madrid, 28006 Madrid, Spain,
Ricardo.Sanz@etsii.upm.es,
WWW home page: http://aslab.disam.upm.es/~rsanz

1 The Stage

Intelligent behavior is more and more required from artificial systems. What was an acceptable behavior for a car control system ten years ago is no longer acceptable for the cars of the immediate future: modern cars do think a lot to economize fuel, avoid collisions and know where they are over the earth. Most machines built today are loaded with embedded intelligence by means of sophisticated control systems.

But control systems engineering is confronting a tremendous challenge as control systems grow in complexity. This challenge can be summarized in three questions:

- What is the best *scientific theory* to support complex control systems engineering processes ?
- What are those *engineering processes* that can guarantee a specified functionality in the final system ?
- What are the *engineering tools* necessary for these engineering activities ?

The theory that traditionally supported control systems engineering was the so called control theory [1], but the central doctrine -as perceived by reading tables of contents from flagship journals- is painfully lagging behind recent advances in computer science, artificial intelligence or robotics.

We can describe the present state of affairs -the need for a new control theory- as the critical need of a *sound theory of mind*. Because what control engineers do is to build *artificial minds* for machines. Human minds are control systems [2] that generate our behavior and control engineers put just simple behavior engines inside artifacts to achieve desired behaviors.

In some sense, the so-called *intelligent control*[3] field emerged some time ago with this objective in mind, but it soon went far from the central first challenge (*i.e.* what to do ?) and got lost into the fields of the second challenge (*i.e.* how to do it?) and particularly of the third challenge (*i.e.* using what ?).

The sound theory is mind is badly needed not only in the control engineering field but in other fields, some of them directly related with the formulation of this theory. Examples are psychology, ethology, epistemology, psychiatry or cognitive science.

A theory able to support engineering processes of mind construction will be also good enough to serve as explanation of biological minds and hence, we can have as a collateral objective for our theory to be good enough to cover these other scientific fields.

2 The Plot

The biggest question for the control systems engineer of today is: *What is the design that better provides the required functionality ?*

Required functionality varies from system to system, from application to applications. It can be as simple as plain setpoint control for single variables to autonomous Mars exploration or optimization of whole refineries.

We must take into account that if we're thinking about real applications, not just laboratory experiments, this means that "requirements" include not only functional requirements but also aspects like cost, maintainability or dependability; this last being extremely crucial in autonomous systems (e.g. flight control systems, intensive care units or nuclear reactor protection). A major concern for complex control systems engineering is that systems size grows more than linearly with their functionality and a complexity handling approach is necessary to overcome these difficulties [4].

It is in the context of intelligent control systems where the issue of control system architecture finds its way. System architecture is the most significant factor related with functionality, constructability and process effectiveness. Architecture centric development processes have demonstrated their suitability in addressing the problems raised by complex systems engineering [5].

Architecture guides systems construction. Systems are built using design patterns that guide developers in the achievement of functionality while keeping complexity under control [6]. The basic design pattern in control engineering is the elementary loop of functioning [7] (*i.e.* perceive-think-act). This design pattern serves as reusable design knowledge to build new systems. Other patterns are available to provide specific functions to the intelligent controller: adaptation, learning, fault-tolerance.

Intelligent systems engineering is using terminology taken from life-related sciences or humanities (biology, psychology, etology, etc) to describe these fundamental designs and properties. Some may argue that this is just lack of imagination to devise new terms, but in most of the cases this borrowing is just cleaning of biologizing lint. Words like knowledge, attention, learning, perception or introspection are common today in intelligent systems literature, having such a clearer meaning that some philosophers are claiming that most philosophical problems of the past have just disappeared.

Lets go a little bit further and argue for the need of devising new architectural designs that can provide our systems with the enhanced capabilities that are desirable today but will be mandatory in the near future. Lets talk about artificial consciousness.

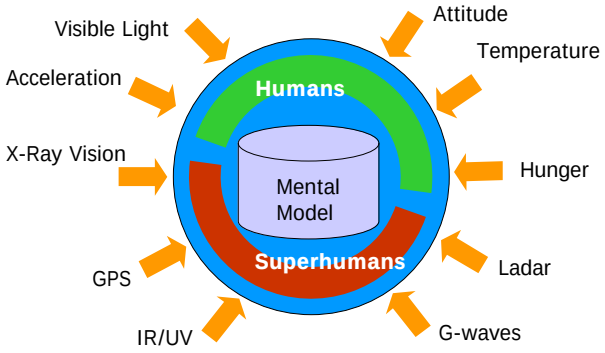

Fig. 1. *Consciousness as world-awareness arises from the run-time integration of perceptual information. In this sense, machines can be quite more conscious of their environments thanks to advanced sensory systems. What machines do lack today is advanced model-based sensor integration mechanisms.*

3 The Central Character

Consciousness is a requirement for top-performing humans in any activity. The hot questions are if it can be also used to describe aspects of artificial systems and if it is at all desirable to have conscious artificialities. While the word actually describes a bag of different concepts like self awareness, world awareness, experience, *etc.*[8]. We will try to argue that some of them are desirable for intelligent control systems.

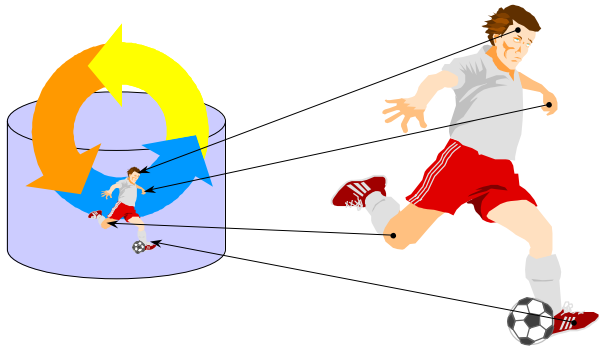

Fig. 2. *Self-consciousness arises when the world model continually updated from the sensor information includes information about the agent itself. Then the agent can reason about its own state and its relation with the environment. In the case of machines this implies and enhanced capability to achieve the objectives specified by its builder.*

The term "consciousness" has no commonly agreed definition but otherwise it has a clear meaning for most of us:

- Searle says: *"... it does not seem to me at all difficult to give a commonsense definition of the term: 'consciousness' refers to those states of sentience and awareness that typically begin when we awake ..."* [9].
- Albus and Meystel say: *"Consciousness is a state or condition in which an intelligent system is aware of itself, its surroundings, its situation, its intentions, and its feelings"* [10].
- Sanz says: *"Consciousness is the process of meaning generation from inputs by a representation centered-system"* [11].
- *et cetera*. See [12], [13], [14] for other interpretations.

The sense of the term that most interest us today is the sense of consciousness as a window to the self. We can argue that consciousness increases as the control system increases complexity (sensors, loops, models, etc) while keeping all its components integrated into a single organization (consider separate the issues of complexity into two drawers: complexity of computations and complexity of the structure of the system i.e. its architecture). The perception of this integrated organization provides the sense of self .

We have raised the issue of maintenance a system-wide unity while dealing with the formidable problems posed to the system. Consciousness is a monitor upon which the self can watch what is going on with this self. We do not equate consciousness and self even while its relation is manifest for many of us (many authors submit to the view that the consciousness is a "looking eye" of our self).

All this becomes extremely important when we start dealing with learning systems of high architectural complexity and continue our pursuit of success by trying to reduce their computational complexity. Knowing about self of these systems can help us in achieving successful functioning. Can we visualize avenues of formal studying self-identity in artificial systems ? Do we have any reason to do this ?

This picture might be meaningful for artificial control systems if we knew for sure that having self is better than having one's personality dispersed over a community of swarm insects. We do not know this for sure, otherwise we would be much more assertive in explaining the precious preferences of having a self, but we have some conjectures.

4 The argument

4.1 Control systems complexity

In an elementary control system, a sensor measure some physical magnitude in the world (for example the speed of a physical machine) and actuators exercise some effect on the world based on the value of the sensor measure 3.

Control systems grow in complexity when required to handle special circumstances. Typically the higher the uncertainty the higher the complexity of

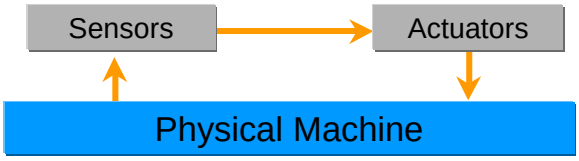

Fig. 3. *An elementary loop of control where action is directly derived from perception. This is sometimes called -in the AI and robotics literature-* reactive control.

the controller. Figure 4 shows a little bit enhanced controller where a filter and an execution monitor are used to reduce the effect of uncertainty (in the sensing process, in the actuation process or even in the setpoint setting process).

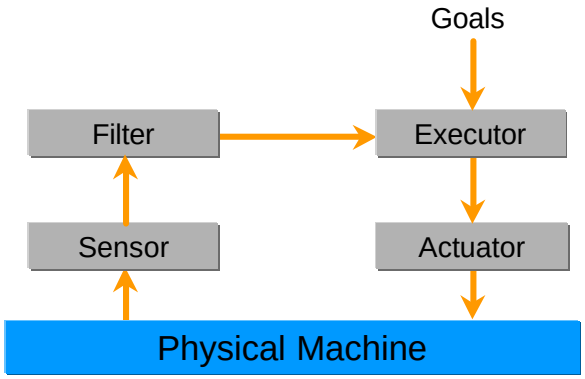

Fig. 4. *Another elementary loop of control enhanced to handle a little amount of uncertainty by means of a filter for perceptions and an execution monitor to control actuation.*

While all controllers do have an internal state, enhanced controllers do manage an explicit state representation that is used in the control process. Form the point of view that the controller is the mind of the machine, this is the *mental state* of it (See Figure 5).

When these states try to reflect the structure of the world under control (*i.e.* they represent the reality) they receive the name of *world models*. They are not fundamentally different from the state representation of Figure 5 but this name do clearly identify its content, making it meaningful for an external observer. Advanced controllers do update this model, not only in the sense of measuring and updating values of model variables, but changing the inner structure of the model, *i.e.* changing the vision of the reality that the machine has. Learning processes do take account for this work, but the lack of dependability of most learning algorithms do limit the available technologies for this task (See Figure 6).

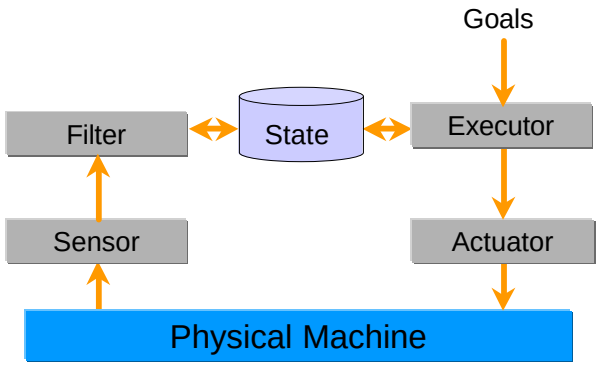

Fig. 5. *State-based controllers do keep explicit representations of controller state that can be assimilated to mental states of the machine.*

4.2 Control with a self

The main contribution from this ”consciousness” perspective to intelligent control systems is perhaps the issue of determining the adequacy of having a uniqueness of self for intelligent systems (i.e. something like an integrated self for an otherwise distributed controller).

The hypothesis is that having self is better to solve control problems with scarce resources in the presence of uncertainty.

Our main conjecture, and it is a central point here, I that the self provides a single, integrated model (with an utility function) for the whole system that can be effectively used to find solutions to the control problems posed continuously to the system.

This enables the global controller to perform system-wide decision making without sacrificing the distributed nature of the control system itself. This distributed nature is necessary to provide effectiveness (e.g. responsiveness, low cost, etc.) and dependability (e.g. tolerating partial system failures, reconfiguration, etc.).

A critical aspect of these ”complex control problems” is the management of uncertainty both external to the agent (i.e. the need of learning) and internal to the agent (i.e. the need of fault tolerance).

Consciousness will be discussed just as a working tool of monitoring processes of functioning not externally, but from inside. Raising the issue of self-identity of learning systems would allow us to pursue all these subjects:

a) without unnecessary extravagance b) attracting people of science not of disturbance c) having continuity with preceding developments.

5 Why do research on artificial consciousness ?

There may be many open questions (Do we have a sound theory of consciousness ?, is artificial consciousness possible?, why do research on it?, how can it

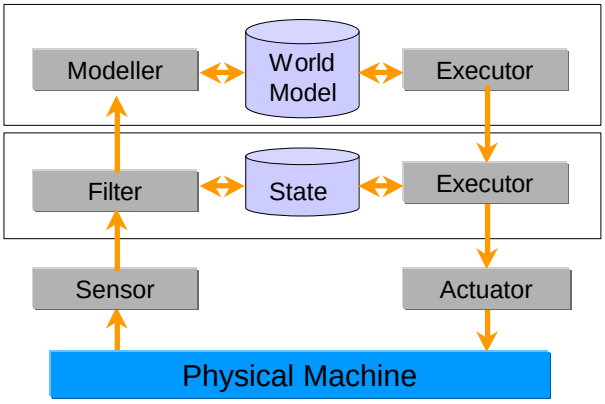

Fig. 6. *Advanced model-based controllers do enhance continuously the explicit representations of reality that the controller uses. This activity is typically called* identification *in control engineering and* learning *in AI.*

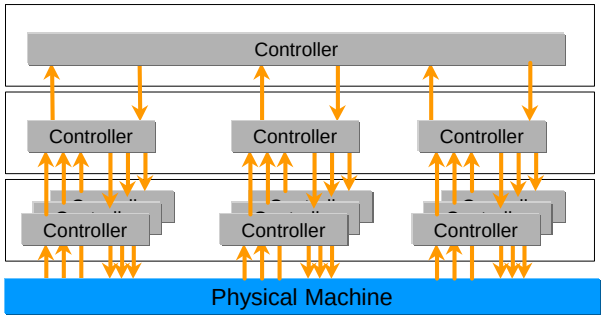

Fig. 7. *Levels of control so perceivable in human minds have a direct correlate with basic hierarchical control systems design in industry.*

be built?, etc.) I just will focus on the "why" issue. There are several reasons why the research on artificial consciousness is important for all of us:

The scientific issue: It will serve as a proof of a cybernetic theory of consciousness, providing an satisfactory scientific explanation of the nature of consciousness. A real conscious machine -just an implementation of the theory- will contribute to solve the sterile debate between hard scientists and mysterians showing how consciousness can emerge just from bare metal. The so called "hard problem" will be shown void of real content .

The technical issue: in the sense proposed by Simon for the term "artificial", i.e. built with a purpose, a conscious control system should perform quite better than an unconscious one (at least this is what we can infer from the natural kinds of controllers: conscious humans perform better than unconscious ones). This is no sense if we are unable to define the terms "conscious

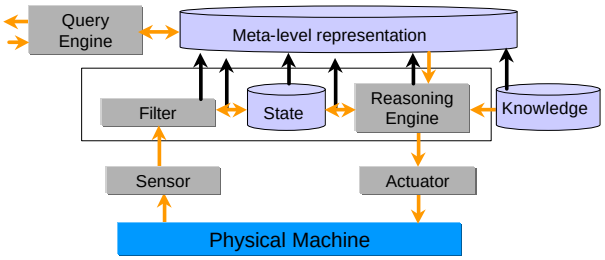

Fig. 8. *Modern control systems do employ meta level representations of themselves to be able to answer queries about their state. This means that present-day advanced controllers do have representations of themselves and hence are self-aware.*

control system" and "unconscious control system", this will be clarified in the central thesis shown later. This conscious behavior is badly necessary in some fields of critical autonomous systems (intelligent weapons, flight control, nuclear power, intensive care units, etc.).

The social issue: if our future is to be machine symbionts as many authors suggest, our survival will depend critically on our capacity of building mental models of machine minds (what psychologists call theories of mind) to interact with them. This will be easier if they think, feel and decide using mental architectures similar to ours.

It is necessary to have a coordinated effort in this field, because visions from differing disciplines are necessary to build a solid theory that could support the technical endeavor of building "artificial people" trustable enough to be dependable.

References

1. Kuo, B.: Automatic Control Systems. Prentice-Hall, Upper Saddle River, NJ (1991)
2. Franklin, S.P.: Artificial Minds. MIT Press, Cambridge, MA (1995)
3. Antsaklis, P., Passino, K.E.: An Introduction to Intelligent and Autonomous Control. Kluwer Academic Publishers (1993)
4. Sanz, R., Schaufelberger, W., Pfister, C., de Antonio, A.: Software for complex control systems. In ström, K.A., Albertos, P., Blanke, M., Isidori, A., Schaufelberger, W., Sanz, R., eds.: Control of Complex Systems. Springer (2000)
5. Garlan, D., Perry, D.E.: Introduction to the special issue on software architecture. IEEE Transactions on Software Engineering **21** (1995) 269–274
6. Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., Stal, M.: Pattern Oriented Software Architecture. A System of Patterns. John Wiley and Sons, New York (1996)
7. Meystel, A., Messina, E.: The challenge of intellligent systems. In: Proceedings of ISIC'2000, 15th IEEE International Symposium on Intelligent Control, Rio, Patras, Greece (2000)
8. Chalmers, D.J.: The Conscious Mind. Philosophy of Mind. Oxford University Press, Oxford-New York (1996)

9. Searle, J.R.: The Mystery of Consciousness. Granta Books, Cambridge, MA (1998)
10. Albus, J.S., Meystel, A.M.: Engineering of Mind: An Introduction to the Science of Intelligent Systems. John Wiley & Sons Inc. (2001)
11. Sanz, R.: Cybernetic consciousness. In: Proceedings of Consciousness and its Place in Nature Conference, Skovde, Sweden (2001)
12. Franklin, S.: Modeling consciousness and cognition in software agents. In Taatgen, N., ed.: Proceedings of the International Conference on Cognitive Modeling, Groeningen, NL (2000)
13. Freeman, W.J.: How Brains Make Up their Minds. Weidenfeld and Nicolson, London (1999)
14. Taylor, J.G.: The Race for Consciousness. MIT Press, Cambridge, MA (1999)
15. Mathis, D., Mozer, M.: The computational utility of consciousness. In Tesauro, G., Touretzky, D., Leen, T., eds.: Advances in Neural Information Processing Systems 7. MIT Press, Cambridge, MA (1995) 11–18