THE AUTONOMOUS SYSTEMS LABORATORY
UNIVERSIDAD POLITÉCNICA DE MADRID

Thinking with the body

Towards hierarchical scalable cognition

R. Sanz, J. Gómez, C. Hernández and I. Alarcón

ASLab R-2008-001 v 1.13 Draft
January 30, 2008

1 Introduction

From the initial, old-age considerations on the nature of mind many very different approaches have emerged. Of special relevance have been the so-called dualist approaches where mind and body have had strongly different aspects and even natures. In this dualistic context for understanding thought and life, the problem has always been how these two realms of being —the physical and the mental— could have any collaboration at all. The mind-body problem was served.

The embodied cognition movement —if we may use this expression— tries to reconcile the apparently multiple quality (*duality* & *unity* at the same time sound kind of religious) by means of analysing the ways in which the body may affect cognition: supporting, raising, sustaining, shaping, etc. Too many terms for a relation that for sure will be much more simple than usually thought or described in excessively lengthy texts. There's nothing to reconcile. Mind and body can't be separated because cognitive agents do think with the body.

We can say that the mind-body problem is not a problem of minds and bodies in the world —*i.e.* a *physical* problem— but a just an artificial, conceptual problem in the minds of the philosophers-scientists. The problem is not that minds and bodies are separate entities but that what is separated are the mental concepts using in thinking about them; *i.e.* that most thinkers use to think about minds and bodies as ontologically fully separated entities. But they are not. Minds can be reduced to bodies because they're just processes running on them.

The approach that we propose may be considered very similar to conventional embodiment or may be perceived as completely apart from embodied cognition and full of panpsychism. There are no minds without bodies and there are no bodies without minds —obviously ranging from the maximally complex to the practically inexistent. There are complexity ranges in all dimensions and cognitive agents go from the simple to the complex in both bodily and mental aspects.

2 Separating mind and body

Following the former analysis, we can say that the correct wording is not that *the mind does emerge-from/is-supported-by/is-shaped-by bodily processes* but that indeed, those *bodily processes are the mind itself*. This may sound obvious for the processes going on inside the brain and leads to the simple identification of the mind with whatever the brain does. With "bodily processes" we are not only referring to those processes happening in the brain but, perhaps in the line of artificial life, to all those information-centric processes that constitute the very inner workings of the hierarchical structures of life.

All orbit around organic modularity (Callebaut and Rasskin-Gutman, 2005) and the robust provision of functions by the different subsystems that constitute a living being. Take for example the case of cardiac pace control (Rezek, 1997). The single purpose of the cardiovascular system is the transport of chemicals to be infused into cells across the whole body. If we consider the heart only, its function is the increase of blood pressure to make it circulate against the resistance of the elastic tissues. To provide the necessary cardiac robustness the heart can autonomously control its beating (see Figure 1).

Obviously the autonomy of heart pacing is limited because it must be responsive to pumping needs of other parts of the body —muscles, viscera, etc.— that are transmitted by different kinds of signals coming from different control levels — and eventually from the cortex-level mind. Indeed, the core system-integrated control of heart rate originates in the circulatory centres of the medulla oblongata and pons —in the brainstem. The control signals reach the heart through sympathetic and parasympatetic nerves to affect many aspects of cardiac operation: force of cardiac contraction (inotropism), rate of cardiac relaxation (lusitropism), heart rate (chronotropism) and impulse conduction (dromotropism) (Opie, 2003).

The heart control example help us address the differentiation and merging between two classes of functions happening in biosystems: core functions and control functions. In the case of the heart, its core function is pumping —with adequate volume and pressure— and is performed by cardiac muscle and cardiac valves. The control function regulates pacing and pressure and it is performed by the sinoatrial and atrioventricular nodes, the bundle of His and the Purkinje cells. However, cardiac muscle is very specialised, being the only type of muscle that is myogenic —which means that it can naturally contract and relax without receiving electrical impulses from nerves— i.e. is mostly autonomous, partially incorporating its own

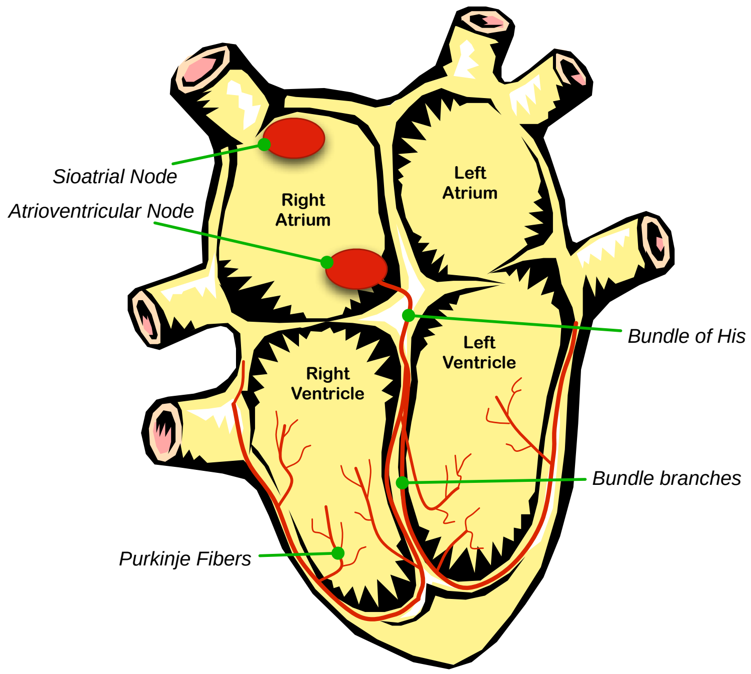

Figure 1: The heart can control its own beating —its core function– almost autonomously. The inherent cardiac rhythm originates at the pacemaker cells of the sinoatrial node to eventually reach the atrioventricular node and through the bundle of His and its three branches -right, left anterior, left posterior- reaches the Purkinje cell network that interacts with myocardic cells. Obviously heart pace can also be controlled by upper control centres in the brain through sympathetic and parasympatetic nerves to be able to respond to demands from higher cognition processes.

control function.

The observation that biological function is performed by a physical core process with an overlaid control process can be generalised down to the molecular biology of the cell and up to the psychophysics of the mind. Control is pervasive in biological function; using the words of Bayliss (1966) we can say that *we are living control systems*. Concrete details of physiological analyses in terms of classical control theory can be found in many places (see for example Khoo (1999)). These analyses also extend to the application of the same control theory of mental-level traits (Powers, 1973; Carver and Scheier, 1981; Nelson, 1993; Marken, 2002).

This dichotomy (core process vs. control process) being pervasive notwithstanding, we should not forget two facts:

- The control processes are necessarily realised as physical process. Information —the very matter of control— does not hold on thin air and needs physical realisations (Landauer, 1992).
- Sometimes the core process and the control process cannot be clearly separated because they are so intermingled in their physical realisation or because the core process has intrinsic control capabilities; *i.e.* it is a self regulated process (as it is the case of myocardic cells thanks to their myogenic capability).

This dichotomy —core processes vs. control processes— is indeed the dichotomy of body and mind. Minds are control processes of physical bodies. But, while this separation can be maximally clear in certain machines that are specifically designed that way (process/control), this is not so in biological systems because evolution tends to favour modularisation of function (Klingenberg, 2005) and not of control and core process; and also *because the very phenomenon of life is in itself a phenomenon of control* over complex physicochemical processes (Rosen, 1991). In any case, minds are informational processes controlling physical processes and realised in them.

From an analytical perspective the mind-body relation is just a relation between the physical and informational realms. The *information extraction frontier* is an interface where the physical reality is transformed into an informational reality. In biosystems this frontier is sometimes not very clear due to the bodily imbrication of control mechanisms but, in general, it is identified as the sensorial system. The *information realisation frontier* is an interface where the informational reality is transformed into a physical reality. Accordingly, this is identified in biosystems with the motor system[1].

[1]Much discussion has been produced around the notion of information in biological systems and the *extraction* vs. *creation* of information by the agent from the environment. Obviously there is information —in the Shannon sense— in the environment. Only part of this is accessible to the agent through its senses and only part is relevant to its dwellings. We agree with the Batesonian view that what is informative for the agent is "difference that makes a difference", *i.e.* information from the world that is profitable for the agent (Bateson, 1979).

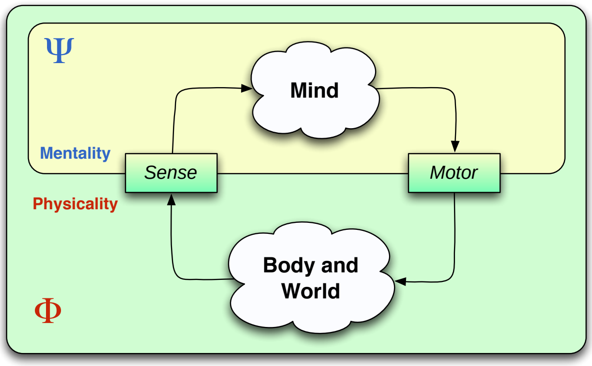

Figure 2: The physical-informational frontier defines the scope of mind .

In technical systems these frontiers are usually well known and engineered for clear separability. It is composed by two sections: sensors and actuators. Sensors map physical into informational. Actuators do the other way, informational into physical.

In a sense, the existence of this frontier is what defines what a mind is. In the purest sense of systems analysis (Klir, 1969) this frontier specifies what should be considered *mental* and what is considered *physical*.

Obviously, in pure psychological domains the firings of Purkinje cells —remember they are muscle cells— are not considered mind at all, but this specification of the information/physical frontier is much more clear that many more arbitrary separations done so far; and is much in line with the separation of what is *system* and what is considered *environment* (see Figure 3).

Additionally, minds are necessarily supported by a physical substrate. The realisation of the informational processes requires the use of physical body parts to handle informational states. These parts can be playing a role in certain core physiological process or can be fundamentally dedicated to informational tasks. This last being the clear case of the CNS. But we should never forget that, in general, the physical coupling of the physical substrate of mind and the rest of bodily processes is maximal — *e.g.* the state and competence of our mind totally depends of blood sugar content in strict connection with liver and pancreas workings.

The need of a physical substrate entails that there are no disembodied minds. What we can find are minds whose physical substrate is partially independent of the physical body they are controlling. Some may understand disembodiment in these terms. In fact this is a basic design objective of control engineers: that the controller substrate is not affected by the physical events in the plant they are controlling. However there is no possibility of total isolation for two reasons:

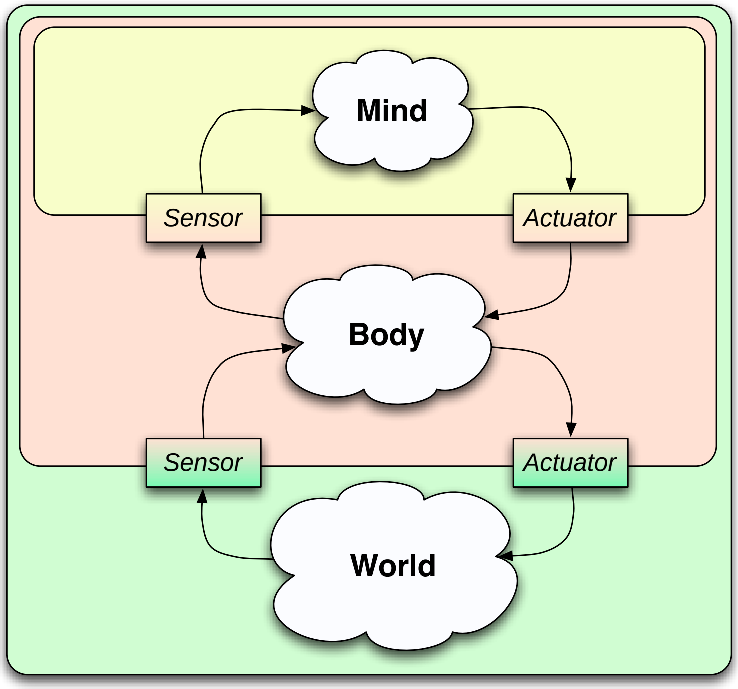

Figure 3: The physical-informational frontier defines the scope of mind and system. But this scope is arbitrary to the extent that —according to systems theory— what is the system and what is the environment is a contingent decision. In the analysis of cognitive systems we use to separate the universe in three arbitrary parts —environment, body and mind— relating states of one to states of another. In some cases the concrete physical realisation of such states is important for the achievement of physical concrete objectives. In other cases it is not.

- Isolation is costly. Replacing copper wires by fibre optics to reduce electromagnetic interference with control signals costs a lot of money and a trade-off is mandatory. So control engineers strive just for sufficiently good isolation. Evolution must have done similarly for biosystems and for the same reasons.
- More importantly, the information frontiers —where the physical becomes mental— are necessarily physical couplings between the mind physical substrate and the rest of the body. While the previous reason was contingent on the economy of the system, this reason for no isolation is absolutely necessary.

So, there are no disembodied minds, unless they are isolated from the body, in which case they are not minds[2]. "Disembodied mind" is an oxymoron. "Embodied mind" is a truism.

3 The phenomenon of control

After the general analysis done in the former section we can start characterising the phenomenon of mind as a phenomenon of control and the proper setting for this analysis is control systems theory and practice. This is a discipline based on dynamical systems theory and specifically focused on this dichotomy body-mind —or plant-controller, as we usually say.

Control theory (Ogata, 1990), as understood in the controls world, is a deeply mathematical theoretical endeavour. Control engineering is the engineering side of it where the theoretical results are put into practice in the form of controllers for machines and processes. But there exists a big gap between theory and practice —as is the case in all engineering disciplines:

- The theoretical results may be of non-applicability for several reasons among which we can mention the following: lack of understanding by practitioners, excessive constraints for their applicability, lack of plants matching the theoretical models, *etc.* .
- There are domains of control technology devoid of theory. This may be due to a lack of interest of control theorists (*e.g.* sensor drift problems) of due to a lack of an adequate formal model (*e.g.* human supervisory control).

In a sense control theory has been driven by its mathematics reaching a situation very similar to that of pure mathematics: disconnection from the real world. And, like most theoretical endeavours, suffers from the "*Consider a spherical cow ...*", syndrome, producing solutions for yet-to-come problems. On the other side, control practice suffers from generalised under-education and a lack of rigour in many of

[2] *c.f.* the well known *brain-in-a-vat* mental experiment.

its activities. This is a purely economical issue, where sufficiently good solutions are enough.

In control systems analysis and design we use the term "plant" to refer to the system we are interested in controlling and "environment" to the rest of the universe (see Figure 4). Obviously, there are interactions between the plant and the environment that affect both the dynamics of the plant and the dynamics of the environment. We are not much interested in isolated systems but as degenerate theoretical cases of this interaction. We can classify the interaction into three classes:

Outputs: The quantities[3] coming from the plant we are interested in. They can be production level in T/h in a cement plant or vehicle speed in a cruise control system.

Inputs: They are the quantities that we can manipulate to drive the plant to the operational point we are interest in. In the former examples, they can be the coal burning rate or the position of the car throttle.

Disturbances: They are material or energetic flows from the environment the we cannot control but nevertheless affect the plant operation. Examples are the level of humidity in raw materials or the wind in the road[4].

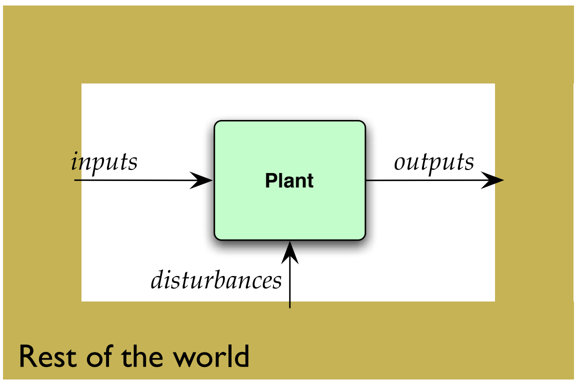

Figure 4: The system immersed in its environment.

The phenomenon of control is simple to state (see Figure 5):

If the dynamics of the interaction of a plant with its environment is not as desired —in terms of some observable quantities—, it is in general possible to

[3]The term *quantity* is used here in the precise sense proposed by Klir (1969).

[4]Strictly speaking there are also undesirable flows from the plant to the environment. In the past they were mostly ignored (if reasonable); now, the widely accepted *ecological* perspective force control engineers to consider them as *outputs to be controlled*.

complement the system with an additional subsystem —let's call it a controller— so that the resultant dynamics of the system plant+environment+controller renders the desired dynamics at the target quantities.

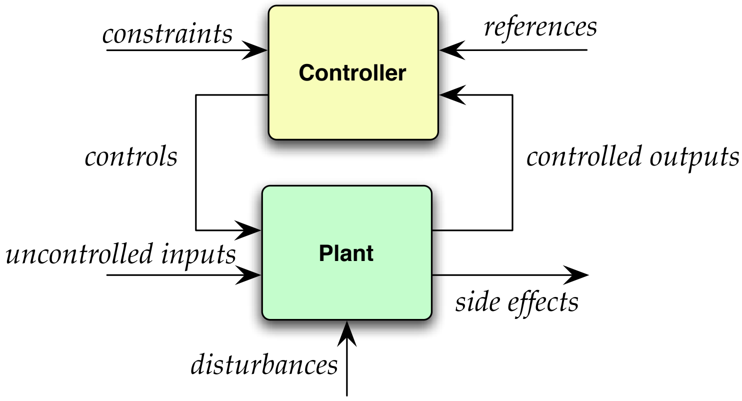

Figure 5: The controller is an additional subsystem so that the resultant dynamics of the system ¡plant+environment+controller¿ renders the desired values at the target output quantities.

The task of devising the proper controller for a given plant and a given set of objectives is called *control design*. In case of biosystems the "designer" is evolution, that is apparently non teleological but if analysed in detail is exactly equivalent to a teleological mechanism addressing *selfish gene* objectives (Dawkins, 1976). The control design problem is an inverse mathematical problem that can be exceedingly difficult to solve (indeed being insolvable in many cases). The common strategy to achieve a solution in difficult cases is dual: try to simplify the mathematical problem doing approximations[5] and relax the target objectives.

We said that it is possible to complement the plant with a controller "in general", because in some cases the necessary controller will be physically irrealizable (*e.g.* would require non-causal behaviour). This comes after the design phase and it is called *controller implementation*. Some of the implementational problems are directly addressed in the design phase (e.g. non-causal controllers are not considered acceptable designs) but other implementational problems can't be handled in the design phase. The principal reason for this is that the construction process cannot be, in general, formalised enough to be of use in the design process that is deeply mathematical.

The most common implementational strategy of today is the construction of controllers in software, as sets of interacting programs. So at the end, the resulting mind controlling the physical body of the plant is a collection of interacting soft-

[5]The most common simplification is to linearise the models of the plant under control (because linear problems are easier to solve).

ware *processes* —this being a precise term in computer science— running atop some computer and communications hardware. The discipline of control engineering has become a discipline of control + computing + communication. The computer metaphor for the mind is no longer a metaphor (Searle, 1990; Cisek, 1999); it is a technological asset.

4 Control from body to mind

A particular topic that is very well addressed by control theory is that of linear feedback control.

A closed-loop feedback controller uses information coming from the actual outputs of the plant to determine the proper control actuations over it. Its name comes from the flow of the information along the system: plant inputs (e.g. throttle) have an effect on the plant outputs (e.g. car speed), which is measured with sensors and processed by the controller; the result (the control signal) is used as input to the plant, closing the action loop (see 6).

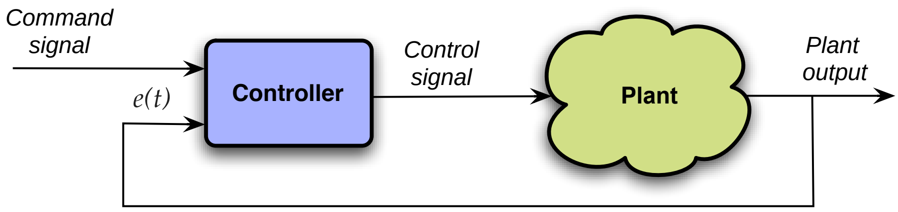

Figure 6: A closed-loop feedback controller uses information coming from the actual outputs of the plant to determine the proper control actuations over it.

In theory, all this complication is unnecessary because if we have a good model of the plant (*i.e.* a formal mapping between its inputs and its outputs) the easiest and perfect control strategy is to invert such model and use it as controller (using the desired output as reference to the controller):

$$desired\ output \rightarrow \mathbf{PlantModel}^{-1} \rightarrow input \rightarrow \mathbf{Plant} \rightarrow desired\ output$$

However there are two big problems in this strategy: i) it is not easy to obtain a perfect model of a plant of non-minimal complexity and ii) this model, if obtained, can be non-invertible. So, in general our controllers are based on plant models that differ from the real plant; this is the reason for using closed loop controllers.

Closed-loop controllers have many advantages over open-loop controllers, mostly related to their capability of handling unmodelled dynamics:

- Disturbance rejection, i.e. make the system robust against unmodelled perturbations (such as unmeasured friction in a motor).
- Performance even with model discrepancies, when the model structure does not match perfectly the real plant.
- Unstable processes can be stabilised, *i.e.* producing a qualitative change in its dynamics (see Figure 5).
- Robustness against plant drift, having reduced sensitivity to plant parameter variations.
- Improved reference tracking performance in the presence of noise.

The major drawback it has is that feedback control is, in a sense, necessarily slower because the controller only reacts when the plant departs from the desired behaviour (*i.e.* after the things went wrong). Alternative structures —like *feedforward* control— are employed to compensate for these drawbacks and not going *behind the plant*.

4.1 PID controllers

The most common control strategy uses a simple linear feedback to compensate errors, speed of change and accumulated error. It is named PID controller —Proportional-Integral-Derivative— referring to the three terms operating on the error signal to produce a control signal (Åstrom and Hagglund, 1995). A PID controller has the general form shown in Figure 7 where $u(t)$ is the control signal sent to the plant by the controller and $e(t)$ is the tracking error $e(t) = r(t) - y(t)$ (where $y(t)$ is the measured output and $r(t)$ is the desired output or reference). K_p, K_d, K_i, T_d and T_i are the adjustment parameters of the controller.

$$\boxed{u(t) = K_p \cdot e(t) + K_d T_d \frac{de(t)}{dt} + \frac{K_i}{T_i} \int e(t) \cdot dt}$$

Figure 7: The PDI controller is the most common control strategy, computing the control signal $u(t)$ from three terms on the tracking error $e(t) = r(t) - y(t)$.

4.2 Mode switching controllers

In general the control capability of a PID controller is sufficiently good to achieve good transitory and steady regime responses. In some cases, however, the simple linear response of the controller (see Figure 7) does not achieve the desired results and alternative strategies must be used. Worst enough, the different strategies will work adequately only in partial regions of the plant state space, hence being no single better alternative to use. In this cases a technique in common use is called mode switching control, where the control system employs one single strategy among several making the choice in terms of the present operational condition of the plant (see Figure 8)

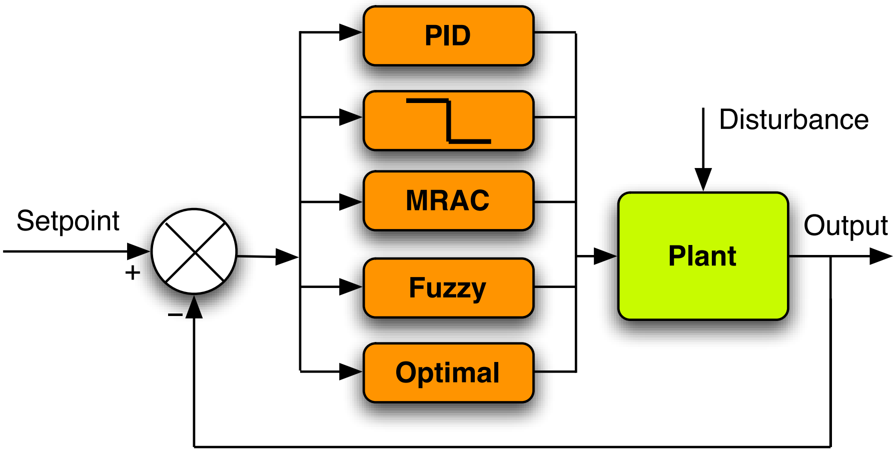

Figure 8: Mode switching controllers (MSC) select the best strategy to use —among a set of predefined ones— for the current situation of the plant. The strategies in this case are i) a PID controller, ii) a bang-bang controller, iii) a model reference adaptive controller (MRAC), iv) a fuzzy controller and v) an optimal controller. References about them can be easily found in the control systems literature (Levine, 1996).

Figure 8 shows a simple structure for mode switching control. The main difference from the basic feedback controller of Figure 5 is the existence of a battery of alternative control strategies:

- a *PID controller*, using the strategy described before;
- a *bang-bang controller*, that provides increased responsiveness sacrificing precision (this is used, for example, in achieving minimum time in subway trains somewhat sacrificing passenger comfort);
- a *model reference adaptive controller (MRAC)*, that will be described later in section 4.4;
- a *fuzzy controller*, that exploits control rules expressed in linguistic terms; and
- an *optimal controller*, that being deeply model-based, tries to maximise certain relevant functional.

However, in the normal use of this control structure, the different controllers employed only differ in their parameterisation and not in their strategy. References about them can be easily found in the control systems literature but a good, all-encompassing text, is that of Levine (1996).

4.3 Model-predictive control

In our former discussion of feedback control we said that the main problem of this strategy is that it is always *behind the plant*, not being able to achieve the desired output values but in response to extant errors.

This is a big problem in certain kinds of systems, where reaching the desired output values in a concrete instant of time is of maximal importance. Think for example of hitting a tennis ball with a racket. The racket must be in the exact position in the precise instant the ball is passing there. Error feedback control cannot achieve this. The only possibility is having some form of anticipation.

Model-predictive control is a strategy based on predicting the future trajectory of a system based on a model of it, and using this anticipatory capability to determine, at present time, the control action necessary for taking the system to a certain state in a precise future instant.

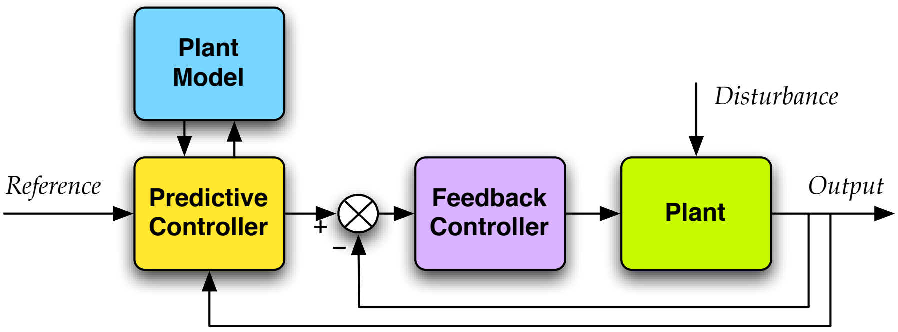

Figure 9: Model-predictive control is a strategy based on predicting the future trajectory of a system based on a model of it, and using this anticipatory capability to determine, at present time, the control action necessary for taking the system to a certain state in a precise future instant.

The realisation of the model predictive controller is usually done in the form of a two-level layered control. The outer layer —the MPC controller itself— uses the plant models and the current plant output measurements to calculate future trajectories of the manipulated variables that will result in operation that fulfils all constraints. The MPC layer then sends this set of manipulated variable changes to the inner layer regulatory controller as setpoints to be pursued in the process.

4.4 Model-reference adaptive control

Adaptive control systems are reflective controllers able to modify its control law to cope with the fact that the parameters of the plant are uncertain or are drifting from the initial conditions. For example, as an spacecraft flies, its mass will continuously decrease as a result of fuel consumption, and the control law must change accordingly. Another example are production processes so harsh —*e.g.* cement manufacturing— that some sensors are continuously changing its measuring capability in a process of continuous grinding.

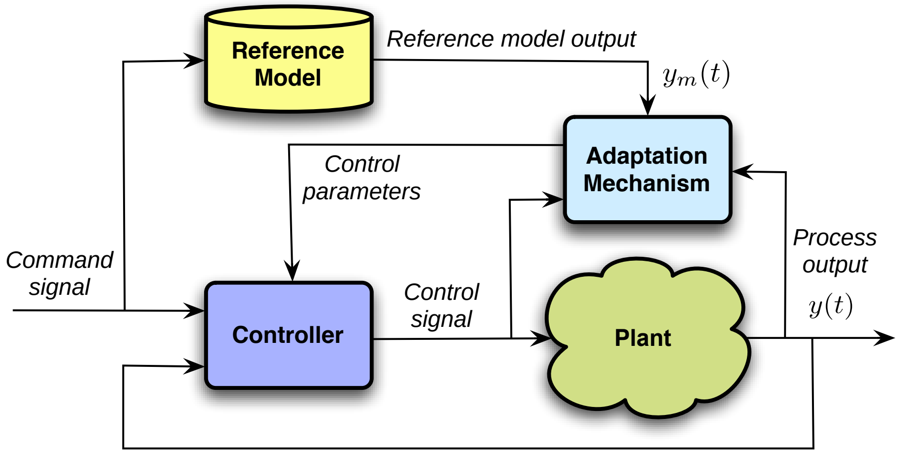

Figure 10: The model-based adaptive controller (MRAC) uses a model of the plant to determine the extent of deviation from the controller design conditions. This deviation is then used to re-tune the controller.

Figure 10 shows the architecture of a model-based adaptive controller. This controller has a control law that is used to control the plant and at the same time uses a model of the plant to determine to what extent the real plant is departing from what was thought. The behavioural differences between the real and the expected are then used by the adaptation mechanism to re-tune the control law parameters to increase it adequacy to the real plant.

4.5 Hierarchical Control

A real plant can be very simple or can be extremely complex. Room thermostats —a favourite in philosophy of mind— are bang-bang controllers of extreme simplicity that are controlling a single magnitude in the plant: room temperature. A real temperature control in a chemical industrial reactor can imply tens of sensors, actuators and heterogeneous nested control loops to achieve the desired performance.

A real industrial plant can have thousands of magnitudes under control and the organisation of all these control loops is a major control system design challenge.

This is so because not only the different magnitudes must be under control but that they must be so in a *co-ordinated way* so as to achieve the global objectives of plant operation.

The strategy to do this is to organise the control loops in a hierarchy where low level references for controllers are computed by upper-layer controllers that try to achieve more abstract and general setpoints. For example, in a reaction unit of a chemical plant many low level controllers control individual temperatures, pressures, flows, *etc.* to fulfil the higher level control objectives of the unit like production and quality levels.

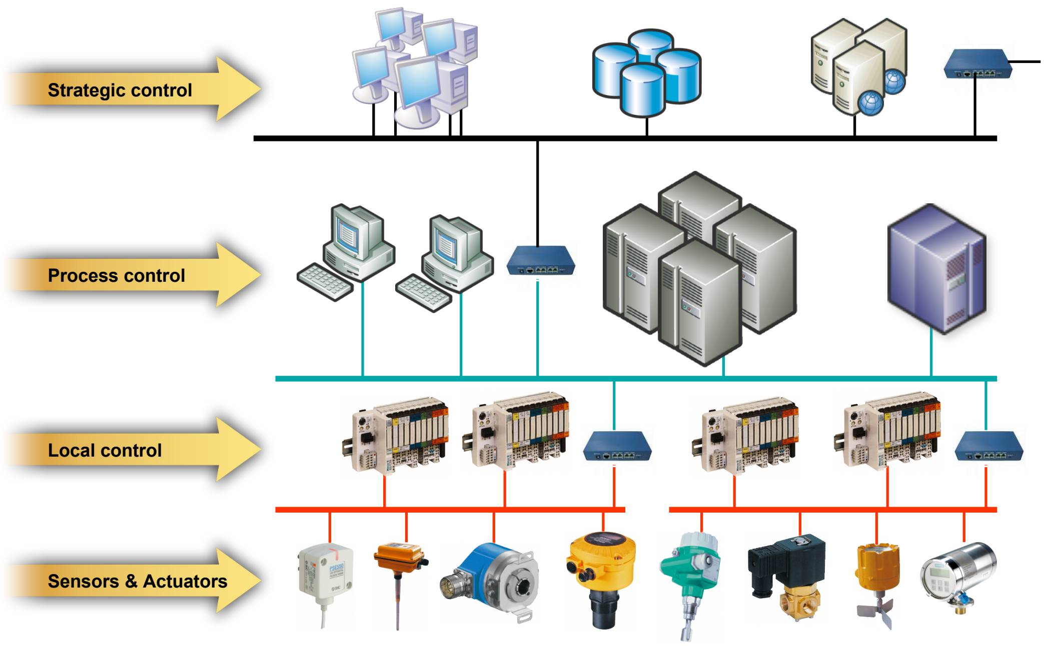

Figure 11: A hierarchical distributed control system (DCS) of an industrial plant is structured in many different control layers. Control objectives gain in abstraction level when going upwards. Temporal criticality and precision grow when going downwards toward the plant.

What is most interesting in studying the phenomenon of control in biological systems and large-scale process plant controllers are the striking similarities between both. While robot control systems in many cases try to mimic biosystems exploiting what is known —or hypothesised— about their control systems, in the case of process plants, the bioinspired movement is yet to come (except possibly at the levels of expert process control (Åström et al., 1986)).

Industrial control systems technology has developed following its own evolutionary path from the early analogical controllers of mid twentieth century to the fully computerised, zillion lines-of-code of today's whole plant controllers. Different ways of organisation have appeared in the structuring of the core processes, in the structuring of control architectures and, quite recently, in the co-structuring of process and control.

From the perspective of the control system we observe an evolution that somewhat parallels the development of mental capabilities in biosystems:

1. The most simple control mechanism is a purely reactive mechanism that triggers some activity when some conditions are met. Examples of these are a big part of all protection and safety mechanisms in industrial systems. The overall behaviour is similar to multitude of safety reflexes in biosystems.

2. An additional level of complexity is achieved when the raw sensorial information is minimally processed to extract meaningful information for the application of behaviour triggering. This is done in elementary control and protection systems. In the case of biological systems, a well known study in this field is the work of Lettwin et al. (1959) about retinal processing in the frog eye[6].

3. The next layer appears when it is possible to conceptualise the operation of the controller and inject on it concrete parametric values (e.g. setpoints or controller parameters). This layer is hence integrable with upper level controls opening the possibility of a control hierarchy. It is also well known in biosystems that some motor actions coming from upstream the CNS are executed by low level controllers (core examples are homeostatic control systems of the body (Cannon, 1932)).

4. Using the conceptual openness of the control loop it is hence possible to layer control loop over control loop —this is called control loop nesting— so that upper level behaviour relies in the robust performance of lower level behaviour —thanks to the integrated controller. In this way it is possible to use a production quality control in a chemical reactor having underneath a plethora of lower level controllers keeping flows, pressures and temperatures at a suitable level. Following the homeostatic example of the previous case, we can discover that large systemic processes —*e.g.* digestion— relay on lower level processes keeping bodily magnitudes apace. Another interesting example is how the process of gait control relies on lower level muscular control (Grillner, 1985).

5. An interesting step forward happens when engineers reach the conclusion that it is possible in general to separate controllers into two parts: an universal

[6] Let's quote this work: "*The output from the retina of the frog is a set of four distributed operations of the visual image. These operations are independent of the level of general illumination and express the image in terms of 1) local sharp edges and contrast, 2) the curvature of edge of a dark object, 3) the movement of edges, and 4) the local dimmings produced by movement or rapid general darkening. [...] Could one better describe a system for detecting an accesible bug?* " (Lettwin et al., 1959).

engine and some data that specifies the concrete control strategy to follow. This opens new possibilities for reuse of engines. A clear example of this are the MPC controllers mentioned in section 4.3 and the controllers based on expert systems technology (Sanz et al., 1991).

6. The next and most interesting step in the development of complex control systems is the realisation that a conceptualisation of this separability engine + knowledge renders a new level of controller openness to metacognitive processes (Meystel and Sanz, 2002). In the case of the human control system this gives rise to introspection capabilities and the the well known phenomenon of memetics and culture (Blackmore, 1999).

What is most interesting of this parallelism between technical industrial control systems and biological controllers is that they have happened in almost complete isolation. For sure, the evolution of technical controllers has not effected substantially the evolution of control mechanisms in biosystems. But the opposite is also true —with the possible exception of knowledge-based control where human expertise was injected into the technical system.

This could be interpreted in the sense that evolutionary pressure on control/cognition points into the direction of layered metacognitive controllers; *i.e.* it points into the direction of consciosness (Sanz et al., 2002). To properly understand this phenomenon let's deepen a bit the analysis of the model-based nature of the control capability.

5 Model-based cognition

As we have seen, models of the plant are critical assets in the construction of control systems. They are used to capture plant dynamics and serve as base information for the control design process. They indeed are explicitly used in some control schemata (*e.g.* model predictive control).

So, controller quality depends heavily on our capability to adequately model the plant to be controlled. But modelling is not only essential as a supporting activity for humans performing a control system design process[7]. Plant models end —in one form or another— as parts of the controller. A controller is as good as is its capability to exploit internalised models of the system it is controlling (Conant and Ashby, 1970). Let's repeat this because it is so important:

> *A controller is as good as is its capability to exploit internalised models of the plant it is controlling.*

[7] A plant model —*e.g.* written on paper— can be considered an externalisation in the line of the extended mind concept (Clark and Chalmers, 1998).

At the end of the day, no matter what is the controller architecture and the design process, plant models become an integral part of effective controllers. Even the simplest controllers (*e.g.* the PID) do have parameters that indeed capture the plant dynamics (the K_p, K_d, K_i, T_d and T_i of Figure 7).

This affirmation is so important for cognitive science because we will later equate *models* with *knowledge* providing a formal grounding for an epistemological analysis of cognition-in-the-world. We must stress the fact that we are moving the problem of knowledge from a cloudy phylosophical context (Gettier, 1963) into the more precise and technable waters of modeling (Bernard P. Zeigler, 2000).

5.1 Structure of a model-based cognitive agent

This model-based control approach is very much in line with current trends in cognitive science that claim for the understanding of mind in terms of internalised models. This started with Craik (1943) in the forties but gained in audience with the works of Johnson-Laird, Gentner and followers (Johnson-Laird, 1983; Gentner and Stevens, 1983) but still has to be accepted as more than just a metaphor. Minds are model-based systems. According to the theoretical, system-level analysis of Conant and Ashby (1970), there is an evolutionary pressure to develop such a kind of architecture of cognition, and cognitive science will, sooner or later, necessarily take the model-based approach as a core disciplinary doctrine.

Figure 12 shows a basic structure for a generalised model based controlled system that can both i) generate its own plant models and ii) accept external injection of explicit knowledge. In the interaction with the plant, the pair action/modeller perform both a change of plant state and a change on the controller itself. The *agent* —let's use this name for the system— is able to create mental representations of the world it is interacting with and that includes both the plant and the world affecting it —its environment.

This kind of agent presents a maximal degree of cognitive autonomy but however cannot escape the constraints imposed by the very realisation of its core processes *Action* and *Modeller*.

Much has been said about processes of mental representation and cognitive dynamics. The model-based control approach to cognition offers, however, a consolidated perspective of the multiple aspects regarding situated, embodied cognition:

The relevance of the body: In section 2 we reached the conclusion that we can't ignore the body as it provides the physical processes that sustain the informational processes of the mind. But beyond that obvious fact, the solution of the control problem in the context of autonomous systems requires an explicit consideration of body dynamics in relation with world dynamics. Realise that this is not an *physicalist* argument but a pure mental argument: the solution of the informational problem requires expressin in bodily terms.

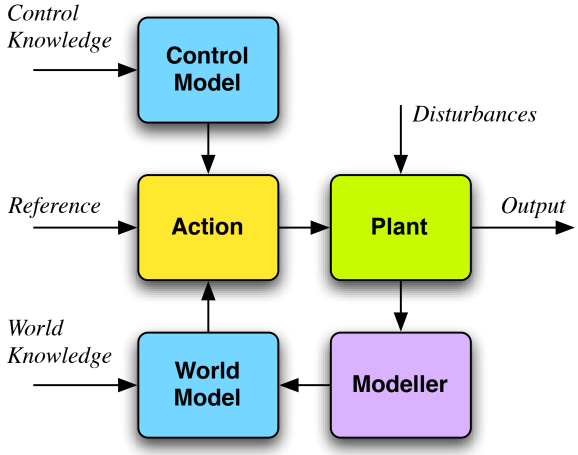

Figure 12: In generalised model-based control, the controller itself is able to generate the world models —a superset of the plant model— to be used in the performance of the control actions. To increase generality the controller employs a source of explicit control knowledge about the control strategy to follow.

The relevance of the task: The age old debate around the possibility of general intelligence somewhat dissipates in this context of model-based control. In Sanz et al. (2000) we analise the problem of *autonomy* in terms of three coupled entities: task, body and world. Solving the control problem requires a maximal precisation of the task; general statements like *achieving human-level problem solving* not only are imprecise but totally misleading because there is no such thing as *human-level competence* except at basic pre-cultural, almost strictly biological capabilities of human bodies. Research on human-level cognitive competences has not properly addressed the question of systemic substrates for the provision of such competences —with some honorable exceptions (Sloman and Chrisley, 2003).

The relevance of the situation: In the very same sense, the dynamics of the reality surrounding the agent must be captured in the mind of the agent. However the necessary deepness of the representation of both agent and environment will be directly dependent on the nature of the control problem to solve and the concrete control strategies followed by the agent. Using robust control schemata (Chen, 2000) the agent will be able to perform sufficiently good without deep representations if body and environment do stay inside some boundary conditions. This is an effective and economic strategy based on maximally simple models (*i.e.* just sets of boundary conditions).

The relevance of epigenesis: The exact amount of epigenesis (Ziemke, 2002) necessary for the agent will depend on many factors: i) the concrete control problem; ii) The amount of *a priori* instantiated knowledge that the agent has in its components; and iii) the level of change in body and environment dynamics that the agent must embrace. So, epigenesis is certainly on the side of adaptation but nevertheless raising a complex mental structure from scratch may take aeons. Bootstrapping will be easier if starting from cognitive engines that are both general and robust. See for example the work of Beer (2003) where a dynamical analysis of a minimal cognitive agent is presented. The agent is implemented by means of a continuous time recurrent neural network that is adaptively configured.

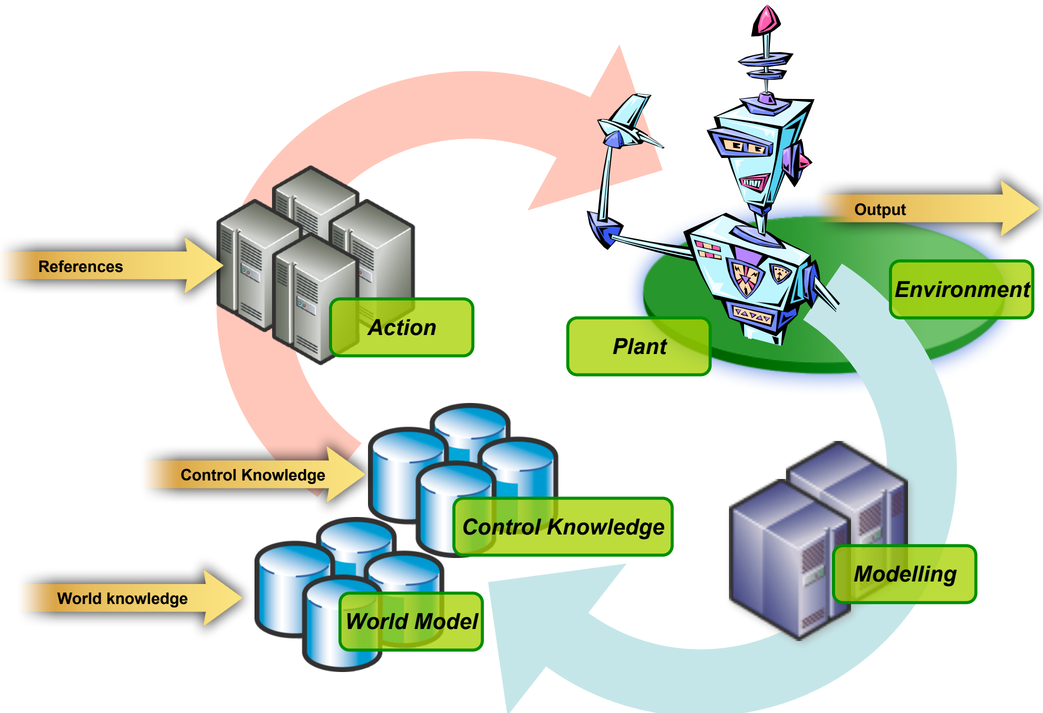

Figure 13: Body-Mind again in a double interaction cycle.

5.2 Operation of a model-based cognitive agent

The global operation of a cognitive system (see Figure 13) is a concurrent activity of two major processes: the process called *action* that maps from the mental to the physical and the process called *perception* that maps from the physical to the mental. The *sense-think-act* classical paradigm of cognition maps into a cycle sense-think-act-behave —much in line with embodied and dynamical systems approaches— that realises the core cognitive pattern (Figure 14).

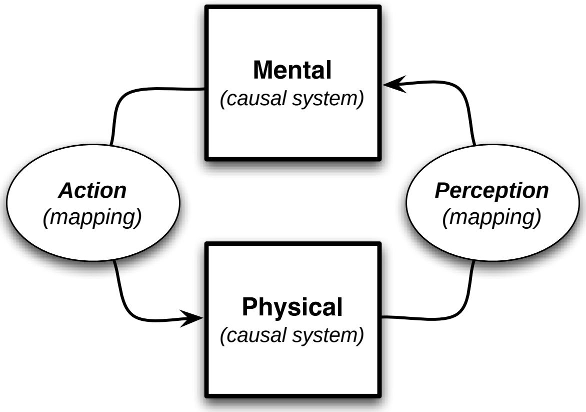

Figure 14: The cognitive pattern.

The operation of a cognitive agent realising the model-based control architecture is very straightforward. In summary:

- The agent extracts relevant information from what Kuipers (2005) describes as *the firehose of experience*. Bear in mind that input happens in two stages, *sensing* and *perception* (López, 2007); the first is constrained by sensor capabilities and the second is limited to what is perceivable in terms of perceptual categories — *i.e.* potential matches in the mental models.

- The different model-integrated percepts change the dynamic equilibrium of mind substrates, settling the operational modes of the agent (Freeman, 2000). This will trigger concurrent mental action dynamics that may produce externalized dynamics —motor outputs— and internalized dynamics —thought.

- The core basic internal dynamical process is the process of mental model construction and maintenance. This happens upon detection of mismatches between perceptual flow and model flow —the product of model execution. The models are based on intrinsic biological capabilities —*e.g.* our genetically implanted model of a human face—, experientially built models —a system identification process (Ljung, 1998)— and culturally transmitted —memetics Aunger (2000).

Notice that we sometimes use *modelling* instead of *perception* to stress the active modelling aspect sublying the perceptual process. In a more detailed architectural analysis these two aspects do have a clear separation of function: perception is external information injection into models and exploits the functionality of the modeller —a model carer.

The approach depicted in figure 12 is maximally general and this implies that in many cases we will find just degenerate versions of it. The most common degenerations are:

1. the collapse of the control knowledge representation and the action generation mechanisms into a single unit (faster, cheaper but less flexible);
2. the elimination of dynamicism in model construction rendering a somewhat classical feedback controller; and
3. the collapse of mental and bodily subsystems in self-regulated processes (*c.f.* section 2).

What is missing to complete the "thinking bodies" picture is the fact that cognitive processes happen at all scales in a complex controlled system. Integration —the key to system level cognition and consciousness— will be devoid of a specific section.

6 Integration is key

Remember that we started this chapter analysing the control capabilities that bodily systems have at different levels. In the case of the heart we saw that these capabilities range from the intrinsic self-organising control capability of myogenic cells to upper brain control.

The structuring of all these control processes is driven by several heterogeneous forces:

- The first one claims for *simplicity* that provides robustness and evolvability.
- The second calls for *non-interference* that enables concurrence and modular evolution of the bodily subsystems.
- At the organism-level efficient behaviour requires that all these separate control loops must be *co-ordinated*.
- *Economic* reasons dictate that many of the functions provided by bodily organs are shared across different subsystems

Integration is the key issue for system-level behaviour (Rossak and Ng, 1991; Grillner et al., 2005). At the end all these forces result in an antithetical decomposition/integration organisation process that renders a unified organism or system.

The concept of system-level integrated functionality is germane to the very concept of biological organism. The idea that different parts of organisms are coordinated to form a functional whole was initially stated by Cuvier as the "principle of correlation" (Cuvier, 1813) and currently it has the denomination of morphological integration (Everett C. Olson, 1999):

> *"This is because the number, direction, and shape of the bones that compose each part of an animal's body are always in a necessary relation to all the other parts, in such a way that - up to a point - one can infer the whole from any one of them and vice versa"*.(Cuvier, 1813)

In computer-based systems the issue of integration has been considered just a matter of proper interfacing in a primary level. However, when addressing issues of large-scale, enterprise-level systems, it is clear that sound integration calls not only for adequate integration mechanisms but also for a unified *integration architecture* (Rossak and Ng, 1991).

The search for such a kind of unified perspective on integrated control architecture in natural systems confronts the pervasive intricateness of biological function and control. However, some interesting works in theoretical biology can be found (Rosen, 1985).

In the case of complex technical control systems (see Figure 15), Sanz (1990) presented an initial layered approach that was refined and extended in following works by Alarcón et al. (1994) and Sanz et al. (1999). We must stress, however, that from a model-based perspective of cognition, integration means control model federation across scales of the control hierarchy[8]. This is an open research issue that is beginning to receive adequate attention (Samad, 1998).

7 Conclusions

Just to conclude this chapter let's summarise its main content in the form of a series of propositions:

- The mind/body relation is an informational/physical relation between a controller and the plant it is controlling.
- The mind —the controller— has a necessarily physical implementation.
- Cognition is the closed dynamical process of sense-think-act-behave. It is a systemic —emergent— phenomenon.
- Cognitive behaviour is based on the exploitation of mental models of plant and environment in the determination of actions.

[8]The very term *cognitive integration* is also used with other integrational meanings, *e.g.* in relation with externalism (Menary, 2007).

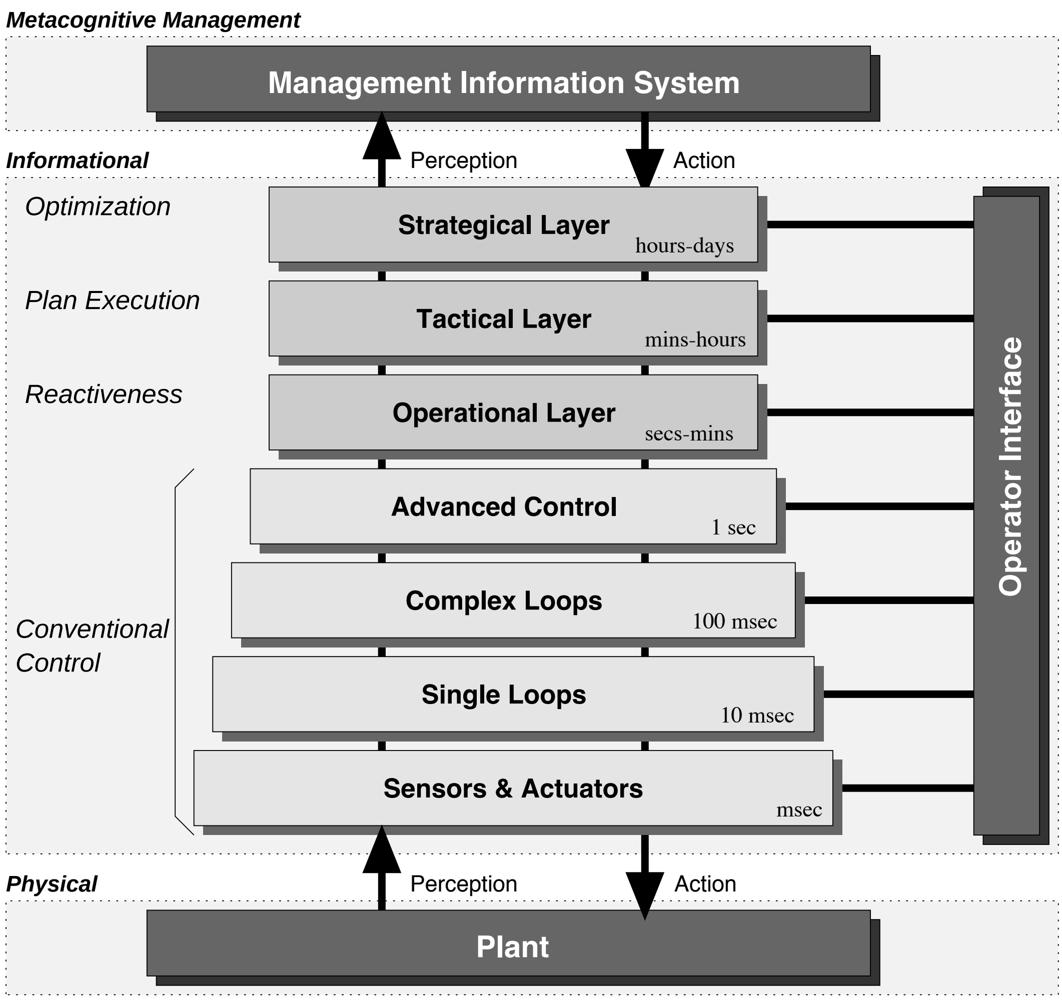

Figure 15: The integration landscape in a whole-plant integrated hierarchical control system. The lower layers provide homeostasis and fast responses to local goal settings. The upper layers implement highly cognitive multi-goal strategic control (including human-in-the-loop integration).

- Sensing is mapping physical states into informational states.
- Actuating is mapping informational states into physical states.
- Perception is model-integration of sensed information.
- Knowledge is executable dynamic models.
- Learning is model creation and caring.
- Cognitive loops organise in heterarchical/hierarchical integrated concurrent systems.

Figure 16 presents a summary depiction of the core fundamental cognitive organisation as derived from the ideas exposed so far. This elementary organisation provides the core cognitive structuring that has to be implemented hierarchically and concurrently in each of the layers of the integrated hierarchy to achieve high-level cognitive capabilities.

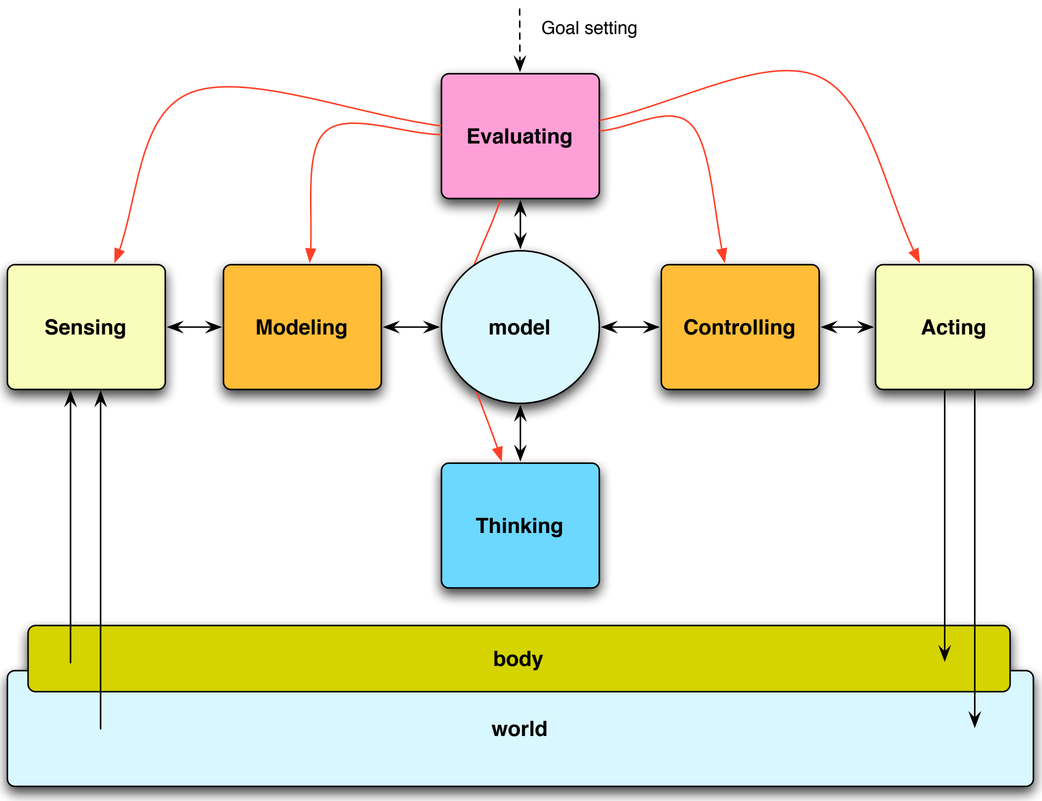

Figure 16: The core epistemic control loop provides the minimal structure for the provision of robust, model-based cognitive capabilities as described in this chapter. This is the minimal building block for general cognitive capabilities.

Cognitive loop integration will render, at the end, a single unified cognitive architecture that can reach, if properly provided with the necessary reflexive self-sensing and self-modelling, any level of cognitive capacity —including self-consciousness (Sanz et al., 2007).

8 Acknowledgements

The research leading to these results has received funding from multiple sources during years but we specifically would like to acknowledge the support in these later stages coming from the European Community's *Seventh Framework Programme FP6/2004-2007* under grant agreement IST 027819 ICEA — *Integrating Cognition, Emotion and Autonomy* and the Spanish *Plan Nacional de I+D* under grant agreement DPI-2006-11798 C3 — *Control Cognitivo Consciente*.

References

Alarcón, M., Rodrí guez, P., Almeida, L., Sanz, R., Fontaine, L., Gómez, P., Alamán, X., Nordin, P., Bejder, H., and de Pablo, E. (1994). Heterogeneous integration architecture for intelligent control. *Intelligent Systems Engineering*.

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence*, 149:91–130.

Åström, K. J., Anton, J. J., and Årzén, K. E. (1986). Expert control. *Automatica*, 22(3):277–286.

Åstrom, K. J. and Hagglund, T. (1995). *PID Controllers: Theory, Design, and Tuning*. The Instrumentation, Systems, and Automation Society (ISA), 2nd edition.

Aunger, R., editor (2000). *Darwinizing Culture: The Status of Memetics as a Science*. Oxford University Press, USA.

Bateson, G. (1979). *Mind and Nature. A Necessary Unity*. Dutton, 2002 edition by hampton press edition.

Bayliss, L. (1966). *Living Control Systems*. The English Universities Press, London.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4).

Bernard P. Zeigler, Tag Gon Kim, H. P. (2000). *Theory of Modeling and Simulation*. Academic Press, 2nd edition.

Blackmore, S. J. (1999). *The Meme Machine*. Oxford University Press.

Callebaut, W. and Rasskin-Gutman, D., editors (2005). *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. The MIT Press, Cambridge, Massachussets.

Cannon, W. B. (1932). *The Wisdom of the Body*. W. W. Norton, New York.

Carver, C. S. and Scheier, M. F. (1981). *Attention and Self-Regulation: a Control-Theory Approach to Human Behavior*. Springer-Verlag, New York.

Chen, B. M. (2000). *Robust and H_∞ Control*. Springer.

Cisek, P. (1999). Beyond the computer metaphor: Behaviour as interaction. *Journal of Consciousness Studies*, 6(11-12):125–142.

Clark, A. and Chalmers, D. J. (1998). The extended mind. *Analysis*, 58:10–23.

Conant, R. C. and Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2):89–97.

Craik, K. J. (1943). *The Nature of Explanation*. Cambridge University Press.

Cuvier, G. (1813). Essay on the theory of the earth.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press, USA.

Everett C. Olson, R. L. M. (1999). *Morphological Integration*. University Of Chicago Press, Chicago.

Freeman, W. J. (2000). The neurodynamics of intentionality in animal brains may provide a basis for constructing devices that are capable of intelligent behavior. In *Performance Metrics for Intelligent Systems (PerMIS)*. NIST.

Gentner, D. and Stevens, A. L., editors (1983). *Mental models*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, (23):121–123.

Grillner, S. (1985). Neurobiological bases of rhythmic motor acts in vertebrates. *Science*, 228(4696):143–149.

Grillner, S., Kozlov, A., and Kotaleski, J. H. (2005). Integrative neuroscience: linking levels of analyses. *Current Opinion in Neurobiology*, 15(5):614–621.

Johnson-Laird, P. (1983). *Mental Models*. Harvard University Press, Cambridge, MA.

Khoo, M. C. K. (1999). *Physiological Control Systems: Analysis, Simulation, and Estimation*. IEEE Press Series on Biomedical Engineering. Wiley-IEEE Press.

Klingenberg, C. P. (2005). Developmental constraints, modules and evolvability. In Hallgrímsson, B. and Hall, B. K., editors, *Variation: A Central Concept in Biology*, pages 219–247. Academic Press, San Diego.

Klir, G. C. (1969). *An approach to General Systems Theory*. Litton Educational Publishing, Inc.

Kuipers, B. (2005). Consciousness: Drinking from the firehose of experience. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 20, pages 1298–1305, Menlo Park, USA. AAAI Press.

Landauer, R. (1992). Information is physical. In *PhysComp '92, Workshop on Physics and Computation*, pages 1–4.

Lettwin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proc. IRE*, 47(11):1940–1951.

Levine, W. S., editor (1996). *The Control Handbook*. CRC-Press.

Ljung, L. (1998). *System Identification: Theory for the User*. Prentice Hall PTR, 2nd edition.

López, I. (2007). *A Framework for Perception in Autonomous Systems*. PhD thesis, Departamento de Automática, Universidad Politécnica de Madrid.

Marken, R. S. (2002). Looking at behavior through control theory glasses. *Review of General Psychology*, 6(3):260–270.

Menary, R. (2007). *Cognitive Integration. Mind and Cognition Unbounded*. Palgrave McMillan, Basingstoke,U.K.

Meystel, A. and Sanz, R. (2002). Self identity in control systems. In *Proceedings of the Workshop on Theoretical Fundamentals of Intelligent Systems*, Durham (NC), USA.

Nelson, T. D. (1993). The hierarchical organization of behavior: a useful feedback model of self-regulation. *Current Directions in Psychological Research*, 2(4):121–126.

Ogata, K. (1990). *Modern Control Engineering*. Prentice-Hall.

Opie, L. H. (2003). *Heart Physiology: From Cell to Circulation*. Lippincott Williams & Wilkins, Philadelphia, 4th edition.

Powers, W. T. (1973). *Behavior: the Control of Perception*. Aldine/Walter de Gruyter, New York.

Rezek, I. A. (1997). *Information Dynamics in Physiological Control Systems*. PhD thesis, Imperial College of Science, Technology and Medicine., London, UK.

Rosen, R. (1985). *Anticipatory Systems*. Pergamon Press.

Rosen, R. (1991). *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. Columbia University Press.

Rossak, W. and Ng, P. (1991). Some thought on systems integration - a conceptual framework. *International Journal on Systems Integration*, 1(1):97–114.

Samad, T. (1998). Multi-modelling: Knowledge integration in complex systems. In Samad, T., editor, *Complexity Management: Multidisciplinary Perspectives on Automation and Control*, pages 55–64. Honeywell Technology Center.

Sanz, R. (1990). *Arquitectura de Control Inteligente de Procesos*. PhD thesis, Universidad Politécnica de Madrid.

Sanz, R., Jiménez, A., and Galán, R. (1991). CONEX: A distributed architecture for intelligent process control. In *Proceedings of the World Congress on Expert Systems*, Orlando, FL.

Sanz, R., López, I., and A.García, C. (2002). On the convergence of natural and artificial consciousness. In *Proceedings of the conference Toward a Science of Consciousness*, Tucson (AZ), USA.

Sanz, R., López, I., and Bermejo-Alonso, J. (2007). A rationale and vision for machine consciousness in complex controllers. In Chella, A. and Manzotti, R., editors, *Artificial Consciousness*. Imprint Academic.

Sanz, R., Matía, F., and Galán, S. (2000). Fridges, elephants and the meaning of autonomy and intelligence. In *IEEE International Symposium on Intelligent Control, ISIC'2000*, Patras, Greece.

Sanz, R., Matía, F., and Puente, E. A. (1999). The ICa approach to intelligent autonomous systems. In Tzafestas, S., editor, *Advances in Autonomous Intelligent Systems*, Microprocessor-Based and Intelligent Systems Engineering, chapter 4, pages 71–92. Kluwer Academic Publishers, Dordretch, NL.

Searle, J. R. (1990). Cognitive science and the computer metaphor. In Göranzon, B. and Florin, M., editors, *Artificial intelligence, culture and language: on education and work*, pages 23–34. Springer-Verlag New York, Inc., New York, NY, USA.

Sloman, A. and Chrisley, R. (2003). Virtual machines and consciousness. *Journal of Consciosness Studies*, 10(4-5):133–172.

Wheeler, T. J. (2007). Analysis, modeling, emergence & integration in complex systems: A modeling and integration framework & system biology: Research article. *Complex.*, 13(1):60–75.

Ziemke, T. (2002). On the epigenesis of meaning in robots and organisms. *Sign Systems Studies*, 30:101–111.

Title: Thinking with the body
Subtitle: Towards hierarchical scalable cognition
Author: R. Sanz, J. Gómez, C. Hernández and I. Alarcón
Date: January 30, 2008
Reference: R-2008-001 v 1.13 Draft
URL: http://www.aslab.org/documents/R-2008-001.pdf

Autonomous Systems Laboratory

UNIVERSIDAD POLITÉCNICA DE MADRID
C/JOSÉ GUTIÉRREZ ABASCAL, 2
MADRID 28006 (SPAIN)