# Chapter 1
# Introduction

## From Brains to the Machines of the Future

**Ricardo Sanz, Carlos Hernández, and Jaime Gómez-Ramirez**

## 1.1 Introduction

Real-world, optimally performant, mission-flexible robots in open-ended environments have been predicted to arrive on a short time by many technologists. Indeed, they have been arriving in 25 years at least during the last 30 years [11].

This is a similar scenario to what has been happening with controlled nuclear fusion. Like fusion reactors, the promised robots are not yet here. The machines of the future are still inside the movies.

Commercial robots—the robots that people will pay-for today—are still only able to operate in controlled or semi-controlled environments doing quite simple tasks: welding car parts in factories or cleaning bathroom floors. The complexities of dwelling in the real world, performing heterogeneous tasks in open-ended, dynamic environments, have proven too difficult for the control technologies available these days.

However, the minute animals in our environment are perfectly able to manage in these conditions. They prove that the problems of real-world activity can be solved in economic ways.

The bio-inspired systems research programme is guided by the idea that the solution can be found in their senses, legs or brains. It should be possible to leverage the technologies produced by Darwinian evolution to improve the behavior of our machines.

## 1.2 Going From Brains to Machines

The behavior of a machine is determined by the interaction between its internal dynamics and the environment it is coupled to. The design process of machines starts in the identification of the desired environmental changes—e.g. moving water from here to there—and goes into the design of a structure for the machine that will bring forth such an effect by the emerging machine-environment interaction dynamics.

When the desired behavior is complex the design of the machine structure is divided into two parts: a physical subsystem capable of producing the environment changes and an informational subsystem

R. Sanz (✉)
Autonomous Systems Laboratory, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain
e-mail: ricardo.sanz@upm.es

C. Hernández et al. (eds.), *From Brains to Systems*, Advances in Experimental Medicine and Biology 718,
DOI 10.1007/978-1-4614-0164-3_1, 

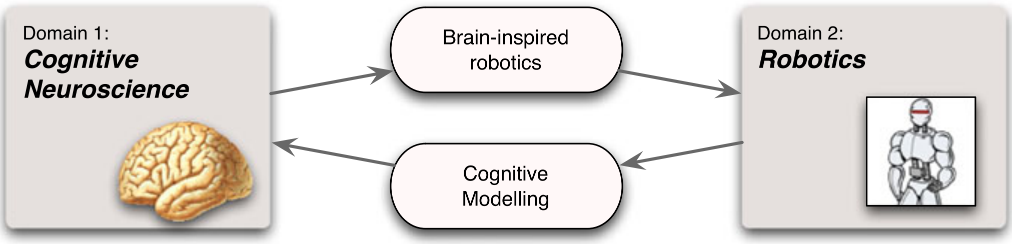


**Fig. 1.1** Brain-inspired robotics serve both as an engineering method and as a experimental testbed for biological theories. Theories of brain function can be tested in robot-based controllers implementing transferred theoretical models

that forces the machine behave in a certain way. The term *controller* is used for the informational subsystem in charge of forcing behavior.

There are different strategies to build controllers. The conventional engineering strategy is to build the controller in such a way that the machine will necessarily behave as desired. Let's call this strategy the *design and build* strategy for artificial minds. This is done using first principles and classic engineering strategies for design [9]. However, when behavior or environment is way too complicated, the design strategy does not cope well.

An alternative strategy is used when the task requirements and its constraints are so complex that we cannot apply the design-build strategy. This strategy is based on reverse-engineering systems that manifest the desired behaviour and copy their functional organisations. This second strategy—let's call it *reverse-engineer and copy*—is what is addressed in this volume, taking the brain as a source of design inspiration.

The study of the brain as source of design knowledge for more effective machines offers the possibility of addressing extremely complex, real-world tasks than only animals can perform so far.

Consider for example the apparently simple task of going around, picking some waste objects to dispose of them into the recycling bin. This is a task that can be done by a toddler or a trained dog. This is also easily done by a robot when the objects and the recycling bin fit certain predefined perceptual categories and the environment conditions are kept in a narrow operative band (e.g. illumination is broadly uniform and sufficient, or the ground is even enough and uncluttered).

However, when these operational conditions are not met, the *designed-built* robot, designed departing from now no longer holding assumptions, will fail in performing the task. Do not try to find articles about robots failing in the scholarly journals of the field. From time to time they appear in funny videos in YouTube but nobody is willing to publish about failures—neither authors nor editors. But failures are there: robots are not functionally robust enough except in performing simple tasks. There is a strong need of improving mission-level robustness if the robots are going to be able to provide their services in open-ended conditions.

It is in this context when we revert to the second strategy: copying brains. Bioinspired cognitive architectures offer the promise of solving this kind of problems because the original architectures—those of brains—are already solving them. There are plenty of threads in this research strategy. Some of them are focused in physical competences of agents but most of them are related to mental competences. Current trends tend to depart from the exploration of the abstractions about the mind and intelligence (as embraced by AI of the sixties), turning to the insights gained exploring the brain [14].

Some will argue that by focusing in the brain we are losing the necessary holistic picture of biological agents. Beyond discussions about embodiment and disembodiment [1, 6] there is a clear need of focusing on the cognitive organisation of the agent (that obviously encompasses the body [12]). While the body is enormously relevant in cognitive processes [2, 10], the role of the brain in higher level cognition is indubitable. The flow from brain knowledge to robotics is a potential source of technological assets. Also, while brain-inspired robotics is a very promising engineering method it is also well settled as an experimental testbed for biological theories (see Fig. 1.1).

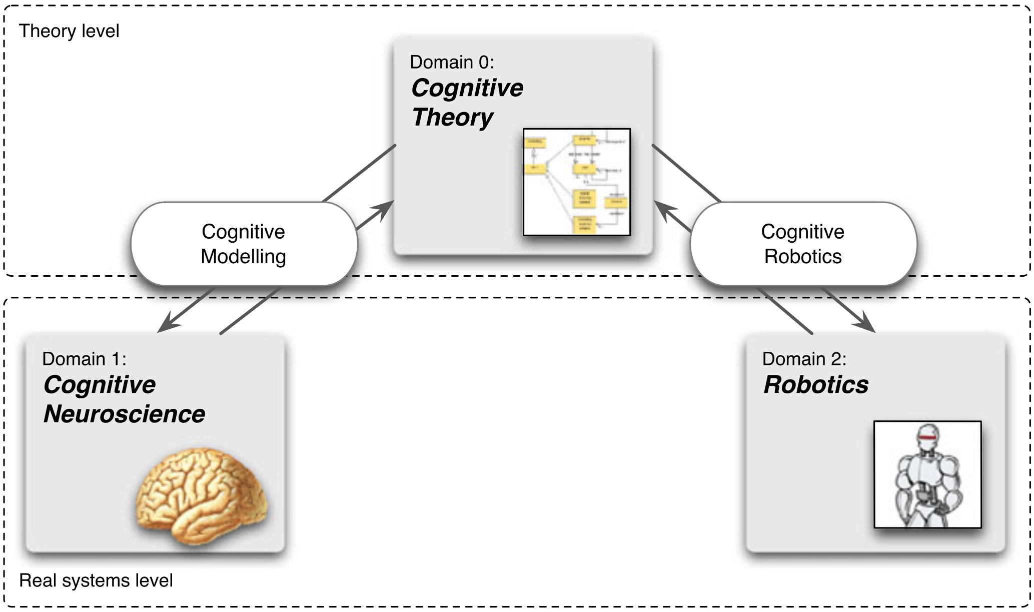


**Fig. 1.2** Brain-inspired robotics will be a technology *sensu stricto* when biological implementation details are abstracted out and only systemic aspects prevail. This implies rendering the theories in cognitive neuroscience in a form that is devoid of biological ties (the abstract *Theory level* shown in the figure above)

Building robot controllers by implementing theories about the brain will serve two basic purposes: (i) controlling the robots and (ii) exploring the implications of the theories and, in a sense, validating them in their ecological contexts [4].

However, systematic cognitive systems engineering requires solid theories and not just a collection of inexplicable designs. It is necessary to transition from a catalog of *ad-hoc* cognitive mechanisms to a rigorous cognitive science; this science may later be applied in the requirements-driven, design process necessary for attaining pre-specified performance levels (see Fig. 1.2).

In bioinspired cognitive systems' engineering it is necessary to extract basic design principles [7, 8]. It is not enough to copy the organisations of animals' brains or bodies [13, 17]. This is the fundamental methodological doctrine behind the several works included in this book: all they try to go beyond the shallow analysis of biological structures, trying to offer more profound, rigorous visions on cognitive systems operation.

## 1.3 Book Contents

The book contains eighteen chapters that cover the whole spectrum of the conference. From models of biological aspects at molecular levels to philosophical considerations about the most abstract aspects of minds.

**Coath et al.**—*Emergent Feature Sensitivity in a Model of the Auditory Thalamocortical System*—investigate plasticity of the brain auditory system. They address the question of whether a recurrently connected thalamocortical model exhibiting spike time dependent plasticity can be tuned to specific features of a stimulus. This is of relevance to the understanding of post-natal—and beyond—construction of cortical and thalamic representations of the features of auditory stimulus that will be available for the cortex-related auditory processing. This work is relevant for the understanding of continuous, post-developmental plasticity that may be critical for robust autonomous systems in changing environments.

**Humble et al.**—*STDP Pattern Onset Learning Depends on Background Activity*—study to what extent the well-known spike-timing dependent plasticity [5] depends on background activity leading even to instabilities. From their results the authors present preliminary insights into the neuron's encoding of temporal patterns of coincidence of spikes and how the temporal precision of the onset response depends on the background activity.

**Basalyga et al.**—*Emergence of Small-World Structure in Networks of Spiking Neurons Through STDP Plasticity*—investigate how a neural network structure changes under synaptic plasticity. They use complex networks of conductance-based, single-compartment integrate-and-fire excitatory and inhibitory neurons showing that under certain conditions, a nontrivial small-world[1] structure can emerge from a random initial network by learning.

**Lefort et al.**—*Coupling BCM and Neural Fields for the Emergence of Self-organization Consensus*—focus on the integration of multimodal perception. They propose a cortex-inspired models for multi-modality association. The model integrates modality maps using an associative map to raise a consistent multimodal perception of the environment. They couple the BCM learning rule and neural maps to obtain a decentralized and unsupervised self-organization.

**Bhattacharya et al.**—*Alpha and Theta Rhythm Abnormality in Alzheimer's Disease: A Study Using a Computational Model*—address theoretical model construction towards solving clinical issues of disease. Their models—of thalamocortical circuitry which exhibits oscillation within the theta and the alpha bands—are aimed at gaining a better understanding of the neuronal mechanisms underlying EEG band power changes. Their work shows how the change in model oscillatory behaviour is related to changes in the connectivity parameters in the thalamocortical as well as sensory input pathways. This understanding of the mechanics under the disease symptomatology may in the future provide useful biomarkers towards early detection of the Alzheimer's disease and for neuropharmaceutical investigations.

**Raiko and Valpola**—*Oscillatory Neural Network for Image Segmentation with Biased Competition for Attention*—study the emergent properties of a cortex-inspired artificial neural network for image segmentation. They combine segmentation by oscillations and biased competition for perceptual processing. They show encouraging results of experiments using artificial image data.

**Johnsson and Gil**—*Internal Simulation of Perceptions and Actions*—address the architectural aspects of neural network architectures based on associative self-organising maps to be able to internally simulate perceptions and actions. They present several topologies—mostly recurrently connected—as e.g. a bimodal perceptual architecture and action neural networks adapted by the delta rule. They show simulation tests that show encouraging experimental results.

**Woodman et al.**—*Building Neurocognitive Networks with a Distributed Functional Architecture*—suggest that the very possibility of successful modeling human behavior with reduced-dimensionality models is a key point in understanding the implementation of cognitive processes in general. They suggest that this is due to a separation in the time scales of the dynamics guiding neural processes and the overall behavioral expression, offering a distributed model based on structured flows on manifolds to understand the organization of this class of behavioral dynamics. They demonstrate this model in a functional architecture of handwriting showing hierarchical sequencing of behavioral processes.

**Schierwagen**—*Reverse Engineering for Biologically Inspired Cognitive Architectures: A Critical Analysis*—analyses methodological and theoretical issues in the development of biologically inspired cognitive systems. He is concerned about the very possibility of reverse-engineering brains by conventional decompositional analysis. Schierwagen concludes that this approach is a no go, discussing the implications for investigations of organisms and behavior as sources of engineering knowledge.

**Quinton et al.**—*Competition in High Dimensional Spaces Using a Sparse Approximation of Neural Fields*—address the computational tractability of implementations of the continuum neural field

[1] A small-world network is a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps.

theory when an adaptive resolution or an arbitrary number of input dimensions is required. They propose a more economic alternative to self-organizing maps using a sparse implementation based on Gaussian mixture models. They test the proposed algorithm in a reactive color tracking application, using spatially distributed color features.

**Aleksander and Gamez**—*Informational Theories of Consciousness: A Review and Extension*—analyse recent theories that establish a systematic link between conscious experience and the flow of information—differentiation and integration—in certain areas of the brain. They analyse measures of information integration [15, 16] and some related algorithms for providing quantitative measures of information integration or causal density; hopefully to be used to make predictions about consciousness. They analyse the computational complexity of these algorithms, which limit their application to just small datasets—networks of around a dozen neurons—implementing one of the better known algorithms in the SpikeStream neural simulator to carry out some experimental comparisons.

**Gómez and Sanz**—*Hippocampal Categories: A Mathematical Foundation for Navigation and Memory*—address the theoretical tools necessary for capturing the theories of cognition that span from neurons to psychological aspects. The mathematical theory of categories is proposed as a valid foundational framework for theoretical modeling in brain sciences, and demonstrated presenting a category-based formal model of grid cells and place cells in hippocampus.

**Dura-Bernal et al.**—*The Role of Feedback in a Hierarchical Model of Object Perception*—address the question of robust object recognition—including occluded and illusory images, or position and size invariances. They propose a model derived from the HMAX model showing how this feedforward system can include feedback, by means of an architecture which reconciles biased competition and predictive coding approaches. This work provides a biologically plausible model of the interaction between top-down global feedback and bottom-up local evidence in the context of hierarchical object perception.

**Manzotti**—*Machine Free Will: Is Free Will a Necessary Ingredient of Machine Consciousness?*—addresses the elusive concept of free will in a mechanistic context. Manzotti analyses whether freedom and consciousness are independent aspects of the human mind or by-product of the same underlying structure; this analysis leads to the author outlining a proposal for an architecture sustaining machine free will.

**Jändel**—*Natural Evolution of Neural Support Vector Machines*—describe two different neural implementations of support vector machines for one-shot trainable pattern recognition. One is based on oscillating associative memory—inspired in the olfactory system—and the second is founded on competitive queuing memory—originally employed for generating motor action sequences in the brain. For both support vector machine models they show that there is a plausible evolutionary path showing that they can apparently emerge by natural processes.

**Chella et al.**—*Self-Conscious Robotic System Design Process—from Analysis to Implementation*—address some of the engineering issues concerning the development of robots endowed with self-conscious capabilities. They analyse the whole engineering lifecycle (from analysis to implementation) focusing on aspects that are specific to the development of self-conscious robotic systems. They propose a new design process—PASSIC—offering custom software engineering techniques for realizing the complex sub-systems needed. This work binds the studies of consciousness with the necessary engineering methods to apply them.

**Arrabales et al.**—*Simulating Visual Qualia in the CERA-CRANIUM Cognitive Architecture*—touch upon the elusive problem of hard consciousness in robots. They attack qualia by a complementary study building “artificial visual qualia” using their cognitive architecture CERA-CRANIUM based on Baars [3] global workspace theory. They study artificial qualia as simulated, synthetic visual experience. The inspection of the dynamics and transient inner states of the cognitive artificial system let them discuss the possible existence of similar mechanisms in human brains.

**Thomsen**—*The Ouroboros Model, Selected Facets*—describes some fundamental aspects of the Ouroboros cognitive architecture: self-referential recursive processes, schema-based memory organisation, feature-driven expectations, etc. Thomsen shows how the Ouroboros Model can address biological cognitive system aspects like attention, emotion, priming, masking, learning, sleep and consciousness.

## 1.4 Value and Perspectives

Science moves in little steps, but also makes its progress with revolutionary discoveries and concepts that sweep away whole and entire edifices of thinking and replace them with new theories that explain more with less. However, there is a constant in this march, the strive for mathematisation and unification.

The extent to which reverse-engineering of brains will help with technological advance in the engineering of more robust autonomous systems is yet to be clear. Nevertheless, the different approaches offered in this book show a steady progress toward more rigorous methods of analysis and synthesis. This rigour implies that they may eventually converge into a single, unified theory of cognition: the very holy grail of cognitive science and engineering.

## References

1. Anderson, M.L.: Embodied cognition: A field guide. Artif. Intell. **149**, 91–130 (2003)
2. Arbib, M.A., Liaw, J.S.: Sensorimotor transformations in the worlds of frogs and robots. Artif. Intell. **72**(1–2), 53–79 (1995)
3. Baars, B.J.: In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. J. Conscious. Stud. **4**, 292–309 (1997)
4. Banerjee, R., Chakrabarti, B.K. (eds.): Models of Brain and Mind—Physical, Computational and Psychological Approaches. Progress in Brain Research, vol. 168. Elsevier, Amsterdam (2008)
5. Bi, G., Poo, M.: Synaptic modification of correlated activity: Hebb's postulate revisited. Annu. Rev. Neurosci. **24**, 139–166 (2001)
6. Chrisley, R.L.: Taking embodiment seriously: nonconceptual content and robotics. In: Android Epistemology, pp. 141–166. MIT Press, Cambridge (1995)
7. Gómez, J., Sanz, R., Hernández, C.: Cognitive ontologies: mapping structure and function of the brain from a systemic view. In: AAAI 2008 Fall Symposium on Biologically Inspired Cognitive Architectures (2008)
8. Hernández, C., López, I., Sanz, R.: The operative mind: a functional, computational and modelling approach to machine consciousness. Int. J. Mach. Conscious. **1**(1), 83–98 (2009)
9. Kuo, B.: Automatic Control Systems. Prentice-Hall, Englewood Cliffs (1991)
10. Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., Pitts, W.: What the frog's eye tells the frog's brain. Proc. Inst. Radio Eng. **47**, 1940–1959 (1959)
11. Moravec, H.P.: Mind Children: The Future of Robot and Human Intelligence. Harvard University Press, Cambridge (1999)
12. Sanz, R., Gómez, J., Hernández, C., Alarcón, I.: Thinking with the body: towards hierarchical scalable cognition. In: Handbook of Cognitive Science: An Embodied Approach. Elsevier, Amsterdam (2008)
13. Sanz, R., Hernández, C., Gómez, J., Bedia, M.G.: Against animats. In: Proceedings of CogSys 2010, 4th International Conference on Cognitive Systems. Zurich, Switzerland (2010)
14. Sanz, R., Meystel, A.: Modeling, self and consciousness: Further perspectives of AI research. In: Proceedings of PerMIS '02, Performance Metrics for Intelligent Systems Workshop. Gaithersburg, MD, USA (2002)
15. Seth, A.K.: Measuring autonomy and emergence via granger causality. Artif. Life **16**, 179–196 (2010)
16. Tononi, G.: Consciousness as integrated information: a provisional manifesto. Biol. Bull. **215**, 216–242 (2008)
17. Webb, B., Consi, T.R. (eds.): Biorobotics. Methods and Applications. MIT Press, Cambridge (2001)