How Information and Embodiment Shape Intelligent Information Processing

Daniel Polani[1], Olaf Sporns[2], and Max Lungarella[3]

[1] Dept. of Computer Science, University of Hertfordshire, Hatfield, UK
[2] Dept. of Psychological and Brain Sciences, Indiana University, Bloomington (IN), USA
[3] Artificial Intelligence Laboratory, University of Zurich, Zurich, Switzerland
d.polani@herts.ac.uk, osporns@indiana.edu, lunga@ifi,unizh.ch

Abstract. Embodied artificial intelligence is based on the notion that cognition and action emerge from interactions between brain, body and environment. This chapter sketches a set of foundational principles that might be useful for understanding the emergence ("discovery") of intelligence in biological and artificial embodied systems. Special emphasis is placed on information as a crucial resource for organisms and on information theory as a promising descriptive and predictive framework linking morphology, perception, action and neural control.

1 Introduction

Artificial Intelligence (AI) strives to understand what "thinking" is by building intelligent entities capable of perceiving, understanding, and manipulating the world surrounding them. The choice of the physical and computational substrates necessary to realize such entities remains a matter of debate. In the early years of electronic computation, one had several different competing approaches to implement processes of thought electronically: cybernetics, systems theory, neural networks, analog machines, and the von Neumann architecture. The classical framework of AI eventually was built on top of the model proposed by von Neumann which emerged as winner. With the success of the von Neumann concept, the algorithmic view of AI prevailed. Intelligence became synonymous with rule-based processing of symbolic information, within a computational architecture that existed independently of its physical implementation. Such a functionalist stance explicitly divorced intelligence from its material or biological substrate. Intelligent systems were targeted at implementing mechanisms derived from the reconstruction of models of human self-inspection or engineered based on technological principles oriented at achieving well-defined and specific goals. In other words, AI became an essentially "platonic" endeavour directed at the top-down design of symbol processing intelligent systems.

While some of the major challenges of AI became reachable (e.g., human-competitive chess-playing software), success was too fragmented. Moreover, there was quite some uncertainty as to what degree one could actually project a phenomenon (e.g., intelligent control) that nature had managed to "engineer" (in fact, evolve) on its own onto a human-designed architecture. Natural solutions have to be *always* viable, i.e., provide stable even if non-optimal solutions in the face of

M. Lungarella et al. (Eds.): 50 Years of AI, Festschrift, LNAI 4850, pp. 99–111, 2007.

uncertainty, noise or incomplete input, or unpredictable changes of context. While viability might seem an incidental property that distinguishes artificial from natural systems, it also fundamentally counteracts the top-down construction philosophy of artificial systems. If this property is taken seriously, emergence of viable solutions for intelligent agents cannot be separated from a permanent entrenchment of the agent in the real world. In particular, the agent's controller is not developed in a platonic world and planted into the agent, but needs to provide the agent with adequate behaviours during its entire lifetime.

Interestingly, a direct danger to the enterprise of platonic "universal" intelligence is posed by the concerns expressed by theorems of the "no free lunch" type [1]. Essentially, such theorems state that finding optima efficiently (and thus efficient learning) is impossible in arbitrary worlds. We do have, however, an existence proof for consistent emergence of intelligence – namely in the biological realm. Biological intelligence appears in as distant species as, say, humans and octopuses; eye evolution reappears in 40-60 different independent lines of descent and often repeats central morphological motifs in remotely related species [2]. In other words, while "no free lunch" type considerations are important for an understanding of abstract "platonic" models, they probably are of lesser relevance for the emergence of intelligence in real-world scenarios. In fact, the world is not arbitrary, but intricately structured and constrained by a subtly intertwined set of properties, e.g., symmetries, continuities, and smoothness. Intelligence is thus *fundamentally* a result of embodied interaction which exploits structure in the world [3]. Two questions remain: how embodiment actually manages to drive the emergence of intelligence under the constraints of uninterrupted viability, as there is no intelligent designer? Is it possible to formulate an architecture-invariant concept that captures the essence of (neural or morphological) computation? If such a concept could be found, one could then apply it to the informational dynamics of an agent acting in its environment. This would yield a computational analogue of what the Carnot-machine is for thermodynamics: by realizing a cycle of information exchange between system and environment, it would provide a consistent framework from which the laws of information processing could be derived given the constraints governing the flow into and out of the system.

Such a perspective would change the way we look at intelligent information processing. Instead of primarily constructing algorithms that solve a particular given task (as in the conventional approach), the phenomenon of intelligent information processing would emerge from an informationally balanced dynamics, without intervention or guidance from an external intelligent designer. Intelligence would be "discovered" rather than engineered (i.e., evoked rather than constructed). For this purpose, it is necessary to identify and formulate suitable quantitative principles. Here, we suggest that Shannon's measure of information [4] (and any quantities derived from it) is a main candidate to help us define such a framework. In the following, we discuss the state-of-the-art of this view on intelligence and how it points towards future perspectives for AI.

2 Information as a Guiding Principle

First attempts to relate information theory to the control of cybernetic systems were done by Ashby [5] who proposed the principle of "requisite variety" (that is, the idea that for control to be possible, the available variety of control actions must be equal or greater than the variety of disturbances). Around the same time, it was suggested that the organization of sensory and perceptual processing might be explained by principles of informational economy (e.g., redundancy reduction [6] and sparse coding [7]). Order and structure (i.e., redundancy) in sensor stimulation were hypothesized to greatly simplify the task of the brain to build up "cognitive" maps or working models of the surrounding environment. In AI, with the increasing dominance of the algorithmic as opposed to the cybernetic approach, the use of information theory was neglected for a long time (applications were typically limited to quantifications of classification strengths, such as in algorithms to generate decision trees). One problem lay in the fact that it was not clear how to make systematic use of information theory to design or understand intelligent systems. In view of the lack of progress, Gibson [8] suspected that the failure of information theory was intrinsic, because in its original form it considers a sender whose intention is to communicate with a receiver, while – so Gibson's argument – the environment of an agent has no intent of informing the agent.

Within the resurgence of neural networks, a major step ahead was taken by Linsker who proposed the principle of maximum information preservation (Infomax; [9]). His objective was to identify principles that would help narrow down architectures plausible for biological settings, e.g., the early processing stages of the visual system. The underlying tenet was the following: due to the intricate processes of the higher-level stages, earlier stages cannot predict what information is going to be used later. The most unbiased hypothesis is thus to assume that earlier stages maximize the total information throughput. In other words, everything that is processed in the later stages of the vision system has to pass through these early stages. This hypothesis was applied to a feed-forward network making some general architectural assumptions, namely a layered structure and localized receptive fields of the neurons arranged in two-dimensional sheets. Maximization of the amount of information preserved at each processing stage caused the neurons' receptive fields to exhibit response profiles similar to the filters found in the vision systems of many organisms. The Infomax principle provides a powerful and general mathematical framework for self-organization of an information processing system that is independent of the rule used for its implementation.

Another dramatic illustration of the central importance of information for living systems comes from work on bioenergetics. Surprisingly, information transmission and processing are metabolically expensive. For example, the blowfly retina consumes 10% of the energy used by the resting fly and, similarly, the human brain accounts for up to 20% of the oxygen consumption at rest [10]. If metabolic cost of a particular informational resource (i.e., neural information processing capacity) is limiting, there is not only a good chance that neural circuits have evolved to reduce its metabolic demands, but also that it will be exploited to a significant degree (and sometimes close to its limit) by a biological system [11]. These results indicate that "information" is almost as important as energy [12]. Motivated by this dominant role of information in living systems, we will therefore suggest to entirely focus on information "metabolism" as *the* single principle driving the emergence and formation of intelligence.

The obvious caveat is that the complexity of living systems may make it hard to pinpoint one single universal principle guiding the emergence of a class of phenomena such as intelligence. Many types of selection pressures, driven by a variety of resource requirements or by other factors (such as sexual selection) act on an organism. Why can we expect that it is sufficient to concentrate on the information "metabolism" to understand the emergence of intelligence? Although present space does not permit a discussion of this question at full length, we wish to reemphasize a few important arguments supporting this view. Not only do sensors and neural structures, as mentioned above, consume a considerable amount of energy in living agents, but also it is known that in living beings sensors and the neural substrate operate close to the theoretically optimal level of information processing. Information is thus a resource of primary importance for a living being and one can expect available capacities to be fully exploited (if not fully exploited, these capacities will degenerate away during evolution). In addition, any further constraints arising from other selection pressures can be factored into the trade-off between available information capacity and the particular interaction dynamics and embodiment.

But perhaps the strongest indicator that universal principles may play a role in the emergence of intelligence is the fact that natural intelligence arises in so many different forms and guises. Species as remotely related and with drastically different sensorimotor and neural substrates as mammals, birds, and octopuses all exhibit a high degree of intelligence. It is hard to believe that evolution would "reinvent the wheel" of intelligent information processing for each of these branches – much more plausible is the assumption of a small number of universal principles giving rise to appropriate evolutionary pressures. Intelligence is, after all, about the right strategy of processing and utilizing information. Therefore, in the following, we will concentrate on the role of *information* in the emergence of intelligence in embodied systems, to the exclusion of any other possible candidate concepts. As we will see, even this restricted set of assumptions provides a rich set of possible paths towards both an understanding of natural as well as the construction of artificially intelligent embodied agents.

3 Information and Embodiment

Once we accept the idea that information is a resource of major importance for living organisms, how does it help direct our attempts to understand the emergence of intelligence in biology and to create intelligence in artificial (embodied) systems?

3.1 Structure of Information

First of all, it turns out that for a living being information is not – as its use as a measure for communication effort might insinuate – a "bulk" quantity. In fact, information is, in general, intricately structured. The information bottleneck formalism [13] illustrates this point most lucidly. Of all the information impinging on a living agent, only a fraction is of true significance to the choice of the agent's actions. This is demonstrated in a striking way by experiments identifying what environmental cues humans are actually conscious of. In controlled settings, for instance, one can show that such cues are surprisingly impoverished (*cf.* phenomenon of change blindness; e.g., [14]). In other words, for a living agent information can be separated into a part that is used (and perhaps recorded), and a part which remains unused.

Such a "split" makes particular sense in the light of the above hypothesis of information being a central resource for living agents. In this case, one would expect evolution to drive up brain power, thus the capacity for information processing until the brain's metabolic costs would outweigh the gains. Thus, a natural limit exists on the amount of information that an agent can process. In other words, information processing must be governed by a *parsimony principle*: only selected components of information should be extracted and processed from the environment, namely those which make the best use of resources with respect to the acquisition and processing cost they entail. Such selected components constitute the *relevant information* available to the agent. It turns out that in typical scenarios, relevant information can be massively reduced while incurring only in moderate losses in overall performance [15]. The performance costs of an agent acting in its environment thus induce structure on information itself by separating relevant from irrelevant information. Information is hence imbued with a "semantic" flavor.[1] But a performance cost profile is not the only factor that provides information with structure. On an even more primordial level, already the *embodiment* of the agent, before the inclusion of any external (evolutionary or task-dependent) performance measures, imposes structure on the *information flows* [3,12,16–19].

To formalize these intuitions, we can express the interaction of the agent with its environment as a causal Bayesian network model of the perception-action loop and consider the information flows arising in the given system [12,20]. The causal Bayesian model allows quantifying the dynamics of the agent as a family of probability distributions, capturing different types of mutual information and information flow quantities in the system. It is found that a given embodiment of an agent favours particular information flows. As a thought experiment, consider, for example, a legged robot where each leg contains a movement sensor. Evidently, one can expect to find the movement sensor mainly reflecting information about the movement of the particular leg it is mounted on, and only to a minor degree that of the other legs. This qualitative intuition can be made precise in a quantitative way and implies the existence of individual, largely separate information flows for the different legs of the robot.

The power of the method extends far beyond this simplistic model and furthermore allows for natural decompositions of information flows. It hence provides a quantitative, theoretically well-founded formalism for characterizing how exactly embodiment induces a bias on *what* information an embodied agent will process and in *which* way. We note that this bias is prior to any concrete goals and tasks the agent may have.

In addition, the "information view" also abstracts away the information processing architecture – which may explain why different species can solve similar tasks using

[1] Semantics was intentionally omitted in the original formulation of information by Shannon, but its absence in the purist interpretation of information theory was long felt to be limiting the potential of information theory to understanding intelligence.

entirely differing brain "hardware." It further relates to the principle of "degeneracy" [21], i.e., the capacity of a system to achieve the same function through many different structural arrangements – a principle found also across individual brains of the same species that are known to differ significantly in terms of their structural components and interconnections, while generating similar perceptual and cognitive performance. Degeneracy fits naturally within an evolutionary framework: as long as brains manage to evolve means to accomplish concrete information processing tasks, it is of minor relevance which part of the brain achieves the task and what its detailed architecture is. The abstractive power of information theory promotes the isolation of necessary from fortuitous aspects of the information processing architecture. It indicates, *ab initio*, what forms of information processing are favoured, prior to any other "implementational" constraints determined by mechanical, biological, and other factors.

Note that on a long time scale, embodiment itself is subject to evolution. Once a concrete embodiment is established, some information flows are reinforced while others are suppressed. It follows that evolution can be seen to operate indirectly on the structure of information flows and even envisage models under which information-theoretic criteria may direct evolution in a Baldwin-like fashion in an environment providing otherwise little short-term fitness structure [22–24]. To simplify the discussion, in what follows, we will restrict ourselves to the case in which the embodiment is given *a priori*.

If the main hypothesis is right that embodiment generates – beyond any concrete implementational substrate – a preferred mode of information processing, then for any given embodied system there should be *natural* controllers. Such controllers would include particular pattern detectors, elementary behaviours and filters, as well as utility (performance) profiles appropriate for the given embodiment [25,26]. The properties of such natural controllers emerge from the *complete* perception-action loop. Note that, at the same time, the time scales characterizing changes of environment, morphology ("hardware"), or controller ("software") are vastly different. Thus, in this picture, the apparent Cartesian duality between body and mind put forward by the classical view evaporates into a mere matter of time scales. In particular, this view suggests that the canonical development of a controller for an embodied system (both in biology and engineering) would first involve starting from the natural information flows emerging from the agent's embodiment, before any concrete tasks and goals are addressed. This is a major deviation from the conventional philosophy which states that the overall control of an embodied agent is attained by a "skillful" combination of partial strategies that achieve individual subgoals. Rather, it makes clear in a mathematically transparent and computationally accessible way how embodiment imposes *a priori* constraints on suitable controllers.

This view provides a plausible argument why nature is able to discover viable solutions for the control of such a variety of "hardware" realizations. It also leads to a novel perspective of how the robot designers could go about designing a controller that is particularly suitable to a given hardware and that could be adapted on-the-fly to any changes of the underlying hardware. Thus, the skeptic's distrust of viewing an agent's life as merely a sequence of goal-driven behaviours maximizing some utility function is vindicated. While a weak notion of "goal" may still exist, the natural way of looking at one's world is prior to all this: it is a basic fact that an agent has a body.

Only then, goals may pop into place. In particular, goals can be shifted adaptively, while the agent is still equipped with a more or less established sensorimotor repertoire.

3.2 Virtue Out of Necessity: Parsimony as Organizational Principle

We start this section by noting that the capacity of perception-action loops to structure information only emerges if the information processing capability is *limited*. Indeed, a system with unlimited information processing capability would have no need to structure information and could process each event as an individual, indivisible, incomparable symbol out of its infinite alphabet of symbols. Thus, it has been proposed that intelligence arises from the need to compress (e.g., sensor data streams can be compressed by identifying regularities and patterns in them; see http://prize.hutter1.net/).

How little information is sometimes necessary to achieve a good chance of survival is exemplified by species of bacteria that can switch randomly from a "safe" hibernating state to a "risky" wake state in which they have the opportunity to reproduce, but are also exposed to possible attacks, e.g., by antibiotics. Recent research indicates that such bacteria do not employ any sensory information to evaluate whether it is safe to switch state, but switch randomly between wake and hibernating states [27]. In information-theoretical treatments of a related scenario (a bacterial colony), it is possible to show that the information processing capacity necessary for survival of an agent can, under certain circumstances, be reduced to zero if fitness requirements are only moderately reduced from the maximum [16]. This observation suggests an *information parsimony* principle: for a given level of required performance, an intelligent agent could aim to minimize its use of information processing resources and the associated expenditures.

A closer look reveals that information parsimony is essentially a "dual" formulation of the Infomax principle: instead of Infomax's view that the agent will maximize the information throughput through a given system, information parsimony emphasizes that, for a given level of performance, the agent will strive to *minimize* the necessary information processing capacity. While both views are mathematically equivalent in the limit of a stationary system, the difference in the formulations essentially emphasizes the time scales of development and evolution. Infomax assumes a given "hardware" for which the throughput is then maximized (i.e., development), while information parsimony assumes a certain performance level (fitness) which needs to be achieved with the least possible informational effort (i.e., evolution).

In both cases, an agent should strive to make use of its informational capacity to its fullest, and use as much as possible of the available information to its advantage; when such information cannot be made use of, evolution will reduce the unused information processing capacity in the long run. One of the most striking examples from biology is the loss of eye function in blind cave animals [28]. In our hypothesis, the parsimony principle will not be limited to this prominent case, but extends to all levels of information processing by the agent. The rigorous formulation of the information parsimony principle lends additional plausibility to approaches to generate intelligent agent controls based on minimal dynamical systems (e.g., [22,29]), and promises additional insights into what lies behind the emergence of apparent intelligence (i.e., organisms doing the right thing) with seemingly very limited abilities.

Additional quantitative principles can be formulated using information-theoretical notions. For instance, the decomposition of information flows (such as the multivariate bottleneck principle [30]) can lead to the emergence of concept detectors (e.g., in a world imbued with a chemical gradient, such decomposition would produce through self-organization detectors for concepts such as direction, parity, long-term or short-term timers [12]). While such low-level concepts seem to emerge in a bottom-up fashion, the question arises in how far the information decomposition view can tackle concept formation in the context of AI. A long-standing challenge, concept formation is being looked at by a large body of work from different viewpoints. Among these, the informational view promises to provide a coherent, far-reaching framework (on a practical level, methods such as independent component analysis or multivariate bottleneck methods are known methods to decompose data into "natural" components [30,31]). In particular, it offers a coherent currency by which the cost of abstraction (relevant to concept formation) can be dealt with universally throughout the system. So, in this view, we propose that concept formation becomes quasi a by-product of the necessity of managing the complexity of processing information. According to this picture, limited informational resources require a decomposition of incoming information flows into largely independent subcomponents of lower complexity which are then handled individually.

The above principles are not arbitrary but arise naturally from the toolbox of information theory. We thus do not just have pockets of isolated quantities, but a whole network of interrelated principles: in fact, information can be seen as forming a language and as such, it allows formulating, quantifying and relating different aspects of information processing in a quantitative, non-metaphorical manner. In summary, the introduction of independent component analysis as a systematic machine learning tool, as well as models such as the regular and multivariate information bottleneck moved us away from the "bulk" picture of information towards a picture of "structured" information in a precise information-theoretical sense. Another important insight is that embodiment is increasingly recognized as a driving factor after decades of an almost Cartesian split in AI research, separating computational processes from a body which was merely regarded as a translation device between the environment and an internal model of the world [22]. One of the most visible expressions of this "paradigm shift" is the increasing interest in morphological computation [3,32,33]. Today, such novel perspectives are joined by a variety of powerful new theoretical and experimental tools which have led to the development of a candidate framework recruiting Shannon information for the study of intelligence (and routes for its emergence) in embodied systems.

4 Outstanding Research Issues

The view that information theory can provide a comprehensive approach for understanding the emergence of intelligence in biological systems and for producing intelligent behavior in artificial systems – while anticipated for decades (e.g., Ashby [5])

– has only recently begun to crystallize. The core issue is to which extent information processing found in biological organisms can be understood in terms of information-theoretical principles. We wish to emphasize once again that information theory does not just apply to high-level overall performance measures of an agent but reaches down to lowest levels of biological information exchange (e.g., the genetic and the neural code [34,35]). We hypothesize that whatever frame is adopted to understand and model biological and artificial information processing, a suitable formalism needs to provide descriptive and predictive power. While a comprehensive (quantitative) framework may remain out of reach for still some time, the adoption of an information-theoretical perspective holds significant promise, as it allows the investigation of a considerable number of relevant issues under a unifying theme. In this section, we would like to discuss several future issues of interest from a higher-level perspective.

One major issue relates to what drives the *evolution* of perception-action loops. We suggest that a major force in their evolution is the selection pressure due to information processing requirements. Going beyond perception-action loops, information theory may even allow the formulation of more general statements about the informational characteristics of evolving systems. For instance, the "principle of ecological balance" [3] (called "complexity monotonicity thesis" in a different context [36]) states that there is a correlation between sensory, motor and information processing capacity. Although a good chance exists that fundamental information-theoretic principles can be identified supporting this thesis, its universality is not entirely obvious. From the observation of living organisms one expects the (potential) sensorial capacity to exceed the motor capacity by far, and memory and the total information processing capacity to be much higher than the sensorial capacity. The question is whether these relations are incidental or universal, and, if the latter is indeed the case, whether this universality can be expressed quantitatively.

Intimately related to this issue is the question of whether concepts such as *relevant information* [15] (i.e., information requirements deriving from external tasks or fitness criteria) could yield a powerful enough drive to instigate an "arms race" between the information necessary to achieve a goal, the information captured by the sensors for this purpose, the required processing capacity of the brain, and the actuator processing capacity necessary to carry out the tasks. This poses central questions concerning the relationship between the co-evolution of brain, morphology, and control and the emergence of complex systems responsive to relevant (structured) information. Complex systems typically contain a high amount of non-repetitive and non-random structure. In particular, the amount of structure of nervous systems can be characterized by a measure of neural complexity which assesses in an information-theoretical context the interplay of highly segregated and highly integrated neural processes [37]. The presumed increase of neural complexity over evolution may simply reflect intense selection pressure on neural structures to efficiently deal with informational challenges posed by co-evolving body morphology, sensorimotor information flows, and eco-niches. As more and more information structure is generated, there is an increased need for neural structures to extract and integrate information, which in turn drives complexity to higher levels.

Two closely related issues are the ones of open-ended evolution and development, that is, how to create systems that do not stop evolving or do not stop learning "interesting" things. There have been various attempts at designing intrinsic motivation systems that capture the drive towards novel and curious situations. They are either based on the notion of *empowerment*, a measure of the power of the agent to modify its environment in a way that is detectable by the agent itself [25], *homeokinesis*, i.e., selection of action sequences which maintain sensorial predictability [24], predictability of action sequences [38], or the maintenance of an abstract cognitive variable, the learning progress, which has to be kept maximal [39]. It is interesting to note that on the physical side, all these approaches seem in one way or another to relate to the notion of maximum entropy production – a principle believed to be relevant in guiding the dynamics in many complex systems (e.g., [40]). In fact, some of the above principles are aptly formulated in the language of information. It thus is natural to explore possible avenues to unify these different but "similar-minded" approaches.

5 Final Remarks

If we, at the end of this chapter, take a step back and, as a final reflection, consider the issues from a bird's eye view, where does this place us? Compared to other sciences, AI is a strangely hybrid creature. For instance, engineering sciences (or more engineering-oriented branches of computer science, such as software engineering) are typically constructive: starting from a more or less uncontested basis, a "code of practice" for the creation of state-of-the-art artifacts is developed. The basis may occasionally be revised to encompass novel developments (e.g., in software engineering: object-orientation or extreme programming), but the task is mostly about improving the paths from a firm basis to an envisaged artifact.

At the other end of the spectrum, we have sciences such as physics which attempt to model observed phenomena. In such sciences, the *foundations*, and not the constructive aspects of a system are at the core of the issues. The physicists aim to simplify and minimize the assumptions behind their models while attempting to capture as many essential features as possible of the examined phenomena. The universal descriptive and predictive power of models such as Maxwell's theory, relativity and quantum theory or thermodynamics are striking and, in fact, one of the mysteries of science.[2]

Biology, on the one hand, resembles physics, as it studies real natural phenomena. On the other hand, it incorporates elements of reverse engineering, as it attempts to disentangle the intricate mechanisms underlying living beings, all of which have historical origins as products of evolution. Unlike physics, however, in biology the reduction of phenomena to few simple principles has until recently only been

[2] Construction was historically not part of the physicist's agenda. This "pure" agenda has begun to change in recent years with the advent of massive computational power, introducing a new branch, *computational physics*, into the picture.

achieved in a few areas (such as the universality of the genetic code); the complexity of biological phenomena, and their place outside of the range of validity of powerful theoretical models such as equilibrium thermodynamics makes it difficult to impose more or less universal organizing principles on the vast collection of empirical data. While evolution provides a universal theory for all living organisms, the nature of this theory is quite different from the sort of theories that serve as the foundations of physics.

AI lies in between physics and biology (in modern AI, biomechanics, material science, neuroscience come also into play and are increasingly superseding the role of psychology and linguistics which dominated classical AI; see [3]). AI belongs to the realm of engineering, and rightly so, because it strives to *construct* intelligent systems. In many aspects, engineering approaches to AI have proven efficient and powerful. However, there is also a universalistic aspiration in AI. Not unlike physics, AI aims to find fundamental principles underlying the emergence of intelligence. This goal is fueled by the observed power of biological systems which achieve intelligence at many different levels, quite unlike the engineered intelligent systems which are usually optimized for one particular task. Biology under the Darwinian stance is "engineered without an engineer", successfully reinventing wheels (eyes, and other "goodies", actually) again and again in an extremely large and complex space – a strong indication that some universal set of principles are at work. A satisfying picture of AI should aim (and hopefully will be able) to isolate and exploit such principles. Consider thus, the grand goal of AI, the one of understanding how intelligence can be "engineered without engineer": it lies between the constructive view of the engineering sciences, the "first principles" view of physics, and the biological view. The latter one is particularly opaque, since any fundamental principles may be buried in volumes of fortuitous historic accidents or restrictions of the biological substrate.

As discussed in this chapter, a primary candidate for building a suitable framework is provided by a suitable adaptation of information theory to the information processing task posed to *embodied* agents: they thus may turn out to serve as the "Carnot-machine" for intelligent information processing. It is striking that information theory which was developed by Shannon essentially as a response to an engineering challenge, not only provides a different way of looking at probability theory (which later was used extensively in AI), but also found to be intimately related to the physical field of thermodynamics. The well-understood formalism of the latter, however, reached its limits in the "exotic" non-equilibrium states of biology. We conjecture that information theory will play an important role in linking the convoluted world of biological information processing, a physics-like set of fundamental principles for intelligent information processing, and the goal of engineering an intelligent system, all in the service of getting closer to the grand vision of artificial intelligence.

Acknowledgements

The authors would like to thank Rolf Pfeifer and the anonymous reviewer for very helpful and clarifying comments.

References

1. Wolpert, D.H., MacReady, W.G.: No free lunch theorems for optimization. IEEE Trans. Evol. Comp. 1, 67–82 (1996)
2. van Salvini-Plawen, L., Mayr, E.: On the evolution of photoreceptors and eyes. In: Hecht, M.K., Steere, W.C., Wallace, B. (eds.) Evol. Biol. 10, 207–26 (1977)
3. Pfeifer, R., Bongard, J.C.: How the Body Shapes the Way We Think. MIT Press, Cambridge, MA (2007)
4. Shannon, C.E.: A mathematical theory of communication. Bell System Tech. J. 27, 379–423 (1948)
5. Ashby, R.W.: Requisite variety and its implications for the control of complex systems. Cybernetica 1, 83–99 (1958)
6. Attneave, F.: Some informational aspects of visual perception. Psychol. Rev. 61, 183–193 (1954)
7. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. In: Rosenblith, W.A. (ed.) Sensory Communication, pp. 217–234. MIT Press, Cambridge, MA (1961)
8. Gibson, J.J.: The ecological approach to visual perception. Houghton Mifflin, Boston (1979)
9. Linsker, R.: Self-organization in a perceptual network. Computer 21(3), 105–117 (1988)
10. Laughlin, S.B., de Ruyter van Steveninck, R.R., Anderson, J.C.: The metabolic cost of neural information. Nat. Neurosci. 1(1), 36–41 (1998)
11. de Ruyter van Steveninck, R.R., Laughlin, S.B.: The rate of information transfer at graded-potential synapses. Nature 279, 642–645 (1996)
12. Klyubin, A., Polani, D., Nehaniv, C.: Representations of space and time in the maximization of information flow in the perception-action loop. Neural Comp. 19(9), 2387–2432 (in press)
13. Tishby, N., Pereira, F.C., Bialek, W.: The information bottleneck method. In: Proc. of 37th Ann. Allerton Conf. on Communication, Control and Computing, pp. 368–377 (1999)
14. O'Regan, K., Alva, N.: A sensorimotor account of vision and visual consciousness. Behav. Brain Sci. 24, 939–1031 (2001)
15. Polani, D., Nehaniv, C., Martinetz, T., Kim, J.T.: Relevant information in optimized persistence vs. progeny strategies. In: Rocha, et al. (eds.) Proc. of 10 th Int. Conf. on Artificial Life, pp. 337–343 (2006)
16. Klyubin, A.S., Polani, D., Nehaniv, C.L.: Organization of the information flow in the perception-action loop of evolved agents. In: Proc. of 2004 NASA/DoD Conf. on Evolvable Hardware, pp. 177–180 (2004)
17. Lungarella, M., Pegors, T., Bulwinkle, D., Sporns, O.: Methods for quantifying the information structure of sensory and motor data. Neuroinformatics 3(3), 243–262 (2005)
18. Lungarella, M., Sporns, O.: Mapping information flow in sensorimotor networks. PLoS Computational Biology 2, 1301–1312 (2006)
19. Ay, N., Polani, D.: Information flows in causal networks. Santa Fe Institute Working Paper 06-05-014 (2006)
20. Klyubin, A.S., Polani, D., Nehaniv, C.L.: All else being equal be empowered. In: Capcarrère, M.S., Freitas, A.A., Bentley, P.J., Johnson, C.G., Timmis, J. (eds.) ECAL 2005. LNCS (LNAI), vol. 3630, pp. 744–753. Springer, Heidelberg (2005)
21. Tononi, G., Sporns, O., Edelman, G.M.: Measures of degeneracy and redundancy in biological networks. Proc. Natl. Acad. Sci. USA, 3257–3262 (1999)

22. Prokopenko, M., Gerasimov, V., Tanev, I.: Evolving spatiotemporal coordination in a modular robotic system. In: Proc. of 9th Int. Conf. on the Simulation of Adaptive Behavior (2006)
23. Sporns, O., Lungarella, M.: Evolving coordinated behavior by maximizing information structure. In: Rocha, et al. (eds.) Proc. of 10th Int. Conf. on Artificial Life, pp. 323–329 (2006)
24. Der, R.: Self-organized acquisition of situated behavior. Theory in Bioscience 120, 1–9 (2001)
25. Klyubin, A.S., Polani, D., Nehaniv, C.L.: Empowerment: A universal agent-critic measure of control. In: Proc. of IEEE Congress of Evolutionary Computation, pp. 128–135 (2005)
26. Tedrake, R., Zhang, T.W., Seung, H.S.: Stochastic policy gradient reinforcement learning on a simple 3D biped. In: Proc. of 10th Int. Conf. on Intelligent Robots and Systems, pp. 3333–3338 (2004)
27. Kussell, E., Leibler, S.: Phenotypic diversity, population growth, and information in fluctuating environments. Science 309, 2075–2078 (2005)
28. Yamamoto, Y., Stock, D.W., Jeffery, W.R.: Hedgehog signaling controls eye degeneration in blind cavefish 431, 844–847 (2004)
29. Beer, R.: The dynamics of active categorical perception in an evolved model agent. Adaptive Behavior 11(4), 209–243 (2003)
30. Friedman, N., Mosenzon, O., Slonim, N., Tishby, N.: Multivariate information bottleneck. In: Proc. of 17th Conf. on Uncertainty in Artificial Intelligence, pp. 152–161. Morgan Kaufmann Publishers, San Francisco (2001)
31. Comon, P.: Independent component analysis. In: Proc. Int. Signal Processing Workshop on Higher-order Statistics, Chamrousse, France, pp. 111–120 (1991)
32. Paul, C., Lungarella, M., Iida, F. (eds.): Morphology, dynamics and control. Special issue of Robotics and Autonomous Systems 54(8), 617–718 (2006)
33. Pfeifer, R., Gomez, G., Iida, F.: Morphological computation for adaptive behavior and cognition. Int. Cong. Series 1291, 22–29 (2006)
34. Adami, C.: Introduction to Artificial Life. Springer, Heidelberg (1998)
35. Zhaoping, L.: Theoretical understanding of the early visual processes by data compression and data selection. Network: Computation in Neural Systems 17(4), 301–334 (2006)
36. Bosse, T., Sharpanskykh, A., Treur, J.: On the complexity monotonicity thesis for environment, behaviour and cognition. In: Int. Conf. on Complex Systems, 1728 (2006)
37. Tononi, G., Sporns, O., And Edelman, G.M.: A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc. Natl. Acad. Sci. USA 91, 5033–5037 (1994)
38. Schmidhuber, J.: Curious model-building control systems. In: Proc. Int. Joint Conf. on Neural Networks pp. 1458–1463 (1991)
39. Oudeyer, P.-Y., Kaplan, F., Hafner, V.V.: Intrinsic motivation systems for autonomous mental development. IEEE Trans. Evol. Comp. 11(1), 265–286 (2006)
40. Dewar, R.: Maximum entropy production and the fluctuation theorem. J. Phys. A: Math. Gen. 38, 371–381 (2005)