Article

Roads to Consciousness: Crucial Steps in Mental Development (Part III)

Uwe Saint-Mont*

Nordhausen University of Applied Sciences, Germany

Abstract

This contribution explains several "roads to self-awareness", all of them based on the natural sciences. The first one follows our bio-psychological evolution. The second road starts with the engineer's point of view and mainly builds on information science and technology, in particular robotics. The third road taken is the most abstract - It exploits complex dynamic systems and their emergent properties. Despite their different origins and methods, these lines of investigation converge. That is, the findings of various fields can be combined into a unified theory of mind and self-awareness, which is the main purpose of this paper. This overall synthesis suggests that the mind results from a multi-hierarchical organizational structure, and self-reflexive flows of information in embodied systems. In addition to this, stable self-awareness appears spontaneously in sufficiently complex robots, when the system's capability of describing itself crosses the level of conceptually clear information processing (thinking). As an application, one obtains a number of construction principles for mentally developing systems that are explained towards the end of this contribution.

Part III of this four-part Article includes: 3. Complex Dynamical Systems.

Keywords: Self-consciousness, self-awareness, free will, dynamic systems, hierarchical systems, language.

3. Complex Dynamical Systems

More is different (Anderson 1972)

There is yet another, more principled and abstract, line of argument leading to consciousness. Apart from putting this remarkable phenomenon in a larger perspective, it gives some concrete ideas about how it develops and how it is structured. Applying these insights may help to program "artificial intelligence".

* Correspondence: Uwe Saint-Mont, Nordhausen University of Applied Sciences, Germany. Email: saint-mont@fh-nordhausen.de

A) Emergence of new properties

For a long time, reductionism ruled. That is, in order to understand some phenomenon, it is crucial to break the phenomenon down into its constituents and analyse their causal relations. Having thus grasped the inner workings of a mechanism and its main elements, one should at least be able to predict its major results. In a sense, this is the overall "modus operandi" of science: analyse an interesting phenomenon in detail, until you have understood what is going on.

However, "the ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. In fact, the more the elementary particle physicists tell us about the nature of fundamental laws, the less relevance they seem to have to the very real problems of the rest of science, much less to those of society. The constructivist hypothesis breaks down when confronted with the twin difficulties of scale and complexity" (Anderson 1972, p. 393). In other words: having understood the details typically does not imply the big picture. Why not?

The reason is that "at each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other… each level can require a whole new conceptual structure. Psychology is not applied biology, nor is biology applied chemistry… the whole becomes not only more than but very different from the sum of its parts" (Anderson 1972, pp. 393-396).

In the last twenty years or so, the associated philosophical position of *emergentism* has gained ground (for a short introduction and a long list of references see de Souza Vieira and El-Hani, 2008), opposing reductionism, and stressing the importance of organization and supervenience. A particularly interesting account of emergent phenomena is given by Deacon (2007). The most important ideas, however, originated in the natural sciences. For an overview see Èrdi (2008) but also Murphy et al. (2007).

Given this viewpoint, one should not expect that self-awareness may be explained by way of reducing it to some fundamental physical law, like Heisenberg's uncertainty principle, or some anatomical detail, like microtubuli (Hameroff and Penrose 2014). On the one hand most scientists would agree that the brain can be reduced to standard physical particles and forces, and that neurons are the basic elements to be considered. (There is no particular mental "stuff", or "vis vitalis", etc.). However, on the other hand, there is also a consensus that the brain's organization - its anatomy and physiology, i.e., its structure and dynamics - is crucial. That is, despite material reduction, it is a very persuasive idea that self-awareness is an emergent

property on a certain level of (biological) evolution, more precisely, of a certain kind of (complex) organization.

One of the main theses of this contribution is the claim that the systematic use of a rich, natural language leads to completely new features, in particular self-awareness. More specifically, the above arguments suggest that our personal self is the consequence of conceptually-clear, circular information processing, enclosing a preeminent mental token for the person processing the information. In a nutshell, the phenomenon of self-awareness is a major consequence of a sophisticated mental organization, i.e., the multi-hierarchical and self-reflexive flow of information in situated robots, based on precise chunks of information about the agent and its environment.

B) Building the final layer

The crucial problem for nature and thus also for engineers and computer scientists consists in constructing a stable hierarchical structure, governing the dynamic flow of information (within the robot, but also in relation to the world outside). Piaget's idea of assimilation and accommodation describes an elementary mechanism, extending the mental system. However, it does not explain how new modules or even layers are created, developed and integrated.

Since we claim that language is crucial for self-consciousness, let us try to explain in some detail how the crucial new feature of language was added, i.e., how the language sub-system may have developed, and how this innovation led to the unique human mind. (Similarly, for every more basic tier, one may describe how the next layer was possibly built on top of the existing structure.)

a. **Reaching the realm of language.** Great apes possess a sophisticated visual image of the world. Of course they are able to hear and to listen, and can produce a broad range of sounds. Linking a specific sound to a particular sensory impression creates a primitive concept. First, perhaps, accidentally, but, if sounds with a particular meaning, words and concepts bring about an evolutionary advantage, it pays to repeat the process of concept-formation. Thus, next to the familiar sensory model of the world they inhabit, a new module begins to develop.

b. **Establishing and consolidating the new function.** Dozens, hundreds, and finally thousands of concepts define a vocabulary which is enriched by every new concept created. Combining these words according to rather constant rules produces an even more powerful way to describe persons, phenomena, and habitat. Thus, with the inner processes of the new module in place, this module consolidates. On the one hand it is thoroughly grounded in bodily experiences (sensory perceptions of all kinds, but also motor actions), on the other hand, it has rules of its own (a grammar). Playing around with concepts and grammar finally creates

the mature functional system of language, i.e., a versatile tool with a complex structure and a "life of its own", able to describe and explain what is going on.

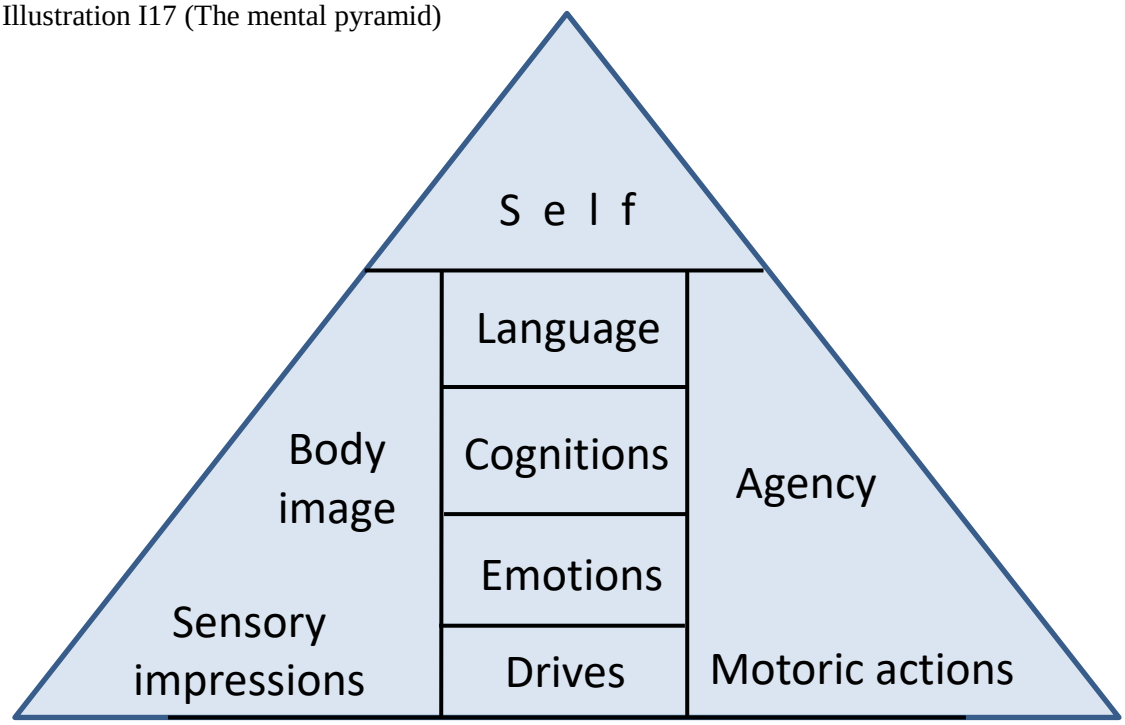
c. **Feedback.** The new functional system, developed alongside the sensor layer, has a major backlash on the received organisation. That is, language and its structure influence sensory impressions or more general cognitions, i.e., the way, we "see" the world. At the very least, concepts provide a second "view", they add precision, and grammatical structures enable arguments and discussions.

d. **The impact of language.** Given the situation of illustration I12 (two interacting functional systems) we moved straightforwardly to illustration I13. In other words, since we are dealing with a recurrent system, we should look for its dynamic equilibrium, i.e., the amount of control exerted. Numerous authors (e.g., Sellars 1970, Campbell 1974, van Gulick 1995, Deacon 1997 & 2007, Murphy et al. 2007) have considered this question. In general, the feedback of a new emergent feature on the elements on which it is founded is called a 2nd order constraint. Such constraints can be very powerful, therefore Haken (2006) uses the term "enslavement," Bunge (2003) speaks of "submergence", and Sperry (1973) coined the term "overpowering". However, with respect to mental processes the most frequently-used term nowadays seems to be "downward determination" or "top-down causation" (or a combination of these words).

e. **Mental reorganization.** All these ideas have in common that they indicate a tremendous influence of language. In effect, with the feedback loops in place, the impact of language leads to a "landslide", altering the mental landscape dramatically. Since we described it in much detail in section 1, a succinct summary should suffice here: First, owing to circularity, there is a particular token standing for one's self on every tier (see Table T2). Second, in particular, there is a word for myself on the top layer, and an elevated image of my body in the visual domain. Third, due to the tight feedback loop between these layers, they act as "internal mirrors". Fourth, the self-concept and the self-image may combine, forming a stable and well-defined (personal) identity that is enriched by contributions of more basic layers (e.g., the protoself). Fifth, drawing a razor-sharp line between ourselves and the rest of the world, we become aware of our precise position in space and time.

Altogether, the edifice looks like a pyramid, governed by a personal self:

Illustration I17 (The mental pyramid)



The *new organization* that has thus evolved features a personal self, i.e., a clear sense of self-awareness, a definite idea of oneself, on top of the restructured mental edifice. This self is rooted in and based on several layers with their particular components (e.g., concepts) and internal structures (e.g., grammar). On the one hand - bottom up - the self incorporates facets of each of the more basic layers (in particular, it has a cognitive and emotional flavour), on the other hand it is an agent, controlling - top down - parts of them. For example, it is able to act (voluntary motor function), talk (conscious command of language), think (intentional use of cognitions), and direct its sensory alignment (focused attention).

However, each layer also has a "life of its own," and the farther away it is from the top, the more so. Owing to the organizational structure, we are able to cope with language best; words are right "on the tips of our tongues", ready-to-use. It is more difficult to control memory and general cognitions: We may not be able to retrieve a certain memory, rotating an object with the inner eye is tedious, and it is very difficult *not* to think of a pink elephant if told so. The motoric realm is divided into voluntary and non-voluntary motor functions. Since we have no direct access to the emotional tier, we are also not able to control emotions directly. If a person is sad, it does not really help to be told to cheer up, and, if frozen in shock, it takes a great deal of voluntary effort to overcome paralysis.

It may be added that quite obviously, the faculty of language, combined with a sense of self, also greatly facilitates and improves communication with others. Thinking in tiers, straightforwardly, a social tier may be added on top of the above pyramid. In other words, thanks to language, the

link between several individuals of the same species is strengthened and more durable social structures than ever before can be built. Thus man became the most eusocial animal ever. With the groups growing in numbers and stabilizing, this layer brought about sedentariness, systematic farming, division of labour, and a multitude of other historical traditions; in the end producing large societies, sophisticated culture and civilization.

Historically, one may distinguish three major shifts: At least 100,000 years ago, spoken language singled out the species of homo sapiens: At about that time proper burials started, a ritual that only makes sense if you have a clear idea about who and where you are (Lieberman 1991). Thousands of years ago, writing greatly increased the ability to store and pass on information, making advanced civilisations possible. The archaeological records for this development are monuments (like pyramids or defensive walls) that could only be erected on the foundations of a sophisticated social organisation. Hundreds of years ago, the formal and quantitative language of mathematics brought about science and technology, i.e., a much deeper understanding and command of all kinds of phenomena, which characterizes the modern era. Thus, taken with a pinch of salt, language–based innovations (printing and the internet included) have exponentiated our ability to learn about nature and ourselves. We truly have become the symbolic species (Deacon 1997), ruling the world.

C) The locus of control[1]

The theory developed above, in particular the last illustration, corresponds nicely to our self-evident "naïve" personal everyday experience. It fits the modern idea of an autonomous agent, but also with the time-honoured view of "free will" (liber arbitirium as it has been called at least since the Middle Ages) that may be traced back as far as Aristotle's "De anima" and Plato's dialogue "Phaedrus" where they depicted the "I" as the charioteer of the soul. Contemporary psychologist Roth (2003) characterizes this idea by saying that the self is in superior command of thinking, planning and action, being a central decision and executive system in the mental realm. He also highlights self-monitoring, self-government and autonomy.

Although the concept of personal freedom has an air of arbitrariness and non-determinism, in particular in the philosophical debate (Kane 2011), it is mostly used in the above sense by natural scientists (Baumeister 2010). So the main connotation of freedom is the self being in charge, having a meaningful choice, being the owner of the mental edifice and the captain of the body. The contemporary challenge to the received top-down conception is bottom-up determinism. The more we have learned about the brain, the more it has become obvious that physical processes

[1] Note that throughout this article, "locus of control" is used in the sense of "who is being in charge", i.e., an entity, that rather issues commands than having to obey them. This is quite different to psychology's use of the term (Rotter 1966).

are the basis of the mental: All thought and emotion, perception and action, memory and personality, depend on anatomical structures and physiological mechanisms. For example, an extraordinary memory (fast, large and reliable) is needed to support language, with the hippocampus (Marr 1971) as well as the classic speech areas of Brocca and Wernicke playing major parts in this narration. In general, brain damage readily implies mental limitations. Thus it is straightforward to conclude that psychology is an epiphenomenon of neurobiology, and the striking well-known results of Libet (1985) and others (in particular, Kornhuber and Deecke 1965) have been interpreted in this way. Given the question "do we consciously cause our actions or do they happen to us?" (Wegner 2002), many natural scientists exploring the brain "bottom up" now opt for the second view. For a thorough discussion see Baer et al. (2008).

However, considering a common computer, it is obvious that problem-oriented programs near the top level drive the physical actions. In the end, the ultimate locus of control is the user, who - via the computer's graphical user interface - tells the software and the underlying hardware what to do. The crucial flow of information consists in commands issued "top down".

How can we explain such a kind of "free will" in self-referential, dynamical systems that we have studied? Close to our account is the idea of a "synergetic computer" (cf. Haken 2004, Haken and Schiepek 2010) which consists of at least two layers - rather organizational than physical - ceaselessly influencing one another.

The first main idea of synergetics is that the elements that make up the bottom layer may interact in a particular way, producing an overall pattern (which, due to its regularity, can be described by a few order parameters), forming the upper layer. The paradigmatic example is light: Unlike ordinary sources of light, a laser does not emit uncorrelated light waves. Instead, it produces a highly coherent single light wave. That's the bottom up impact, i.e., the elements' *spontaneous self-organization* into a larger and simple structure.

The second main idea of synergetics concerns the top-down impact, i.e., the consequences of the large structure (often represented by its order parameters) on the elements in the bottom layer. In the case of the laser, the coherent light wave forces single photons to oscillate in the same way. That is, the elements are no longer "free to do what they like". Instead, they lose many, if not most, degrees of freedom and have to comply with the overall organizational structure (therefore the term "enslavement" mentioned before).

With respect to information flows in the brain, this account captures many important aspects. It is both elegant and there is much experimental evidence in favour of it:

1. We have stressed the importance of feedback, i.e., of an account that is dynamic as well as circular. Synergetics explains how a hierarchical system endowed with a feedback loop/ circular causality may emerge spontaneously via "self-organization".

2. Looking at the lower tier, there are indeed coherent waves when areas of the brain work together (Singer 2007) which are to be expected when parts – via a common order structure – co-operate (automatic "consensus-building" in the words of Haken and Schiepek 2010). More generally, in this view the "binding problem" (how can different brain parts work together when necessary) is solved via spontaneous synchronization bottom up (e.g., Fingelkurts and Fingelkurts 2004, 2013), in particular frequency locking (Haken 2002, 2008).

3. Looking at the upper tier, there is an enormous degree of information compression, since a few order parameters suffice to describe the overall behaviour. It is well known that we do not store all details of a story or picture. Rather, we retain the most interesting, striking and characteristic features.

4. Since patterns may act as the building blocks for further layers, picturing a multi-tier system is straightforward.

5. Information processing within this system is both massively parallel (since there are many modules and feedback loops) and integrated (since the circuits are all interwoven). Moreover, in stark contrast to the classic von Neumann computer architecture, most components are active most of the time.

6. Memory building and pattern recognition use the same mechanism, i.e., the feedback loop between the layers. On the one hand, "bottom up" memory building is self-acting, and to store some pattern it suffices to retain its order parameters. On the other hand, suppose there are stored parameters and some of the pattern's features are observed. The "top down" part of the feedback loop between the layers will then fill in the missing parts, until the dynamics have automatically restored the whole pattern (cf. Haken 2004, Section 17.1). In other words, synergetics offers an elegant, "combined" mechanism of memory building and information retrieval. Data compression and recovery are understood as a kind of feature extraction and pattern formation.

7. More generally, pattern formation is a particular kind of phase transition. Without a pattern, neural activity is incoherent, yet with a pattern it is orderly. Moving to the orderly state involves characteristic fluctuations, critical slowing down, and hysteresis that have all been observed in the motor arena (Haken 2006 (Chapter 11), Haken 2004 (Chapter 12), Haken 1996 (Part II), Haken, Kelso and Bunz 1985).

8. Different patterns are associated with distinct values of the order parameters. Here, too, typical oscillations can be observed, in particular if the sensory input supports several patterns. This effect can be demonstrated nicely with the help of flip-flop images, like Necker's cube (Haken 2004, Chapter 13). Difficult decisions seem to be similar: given a certain information basis, it may be hard to choose between several options, particularly if they are equally promising (Haken 1996, Chapter 17).

9. It is well known that layers building on each other operate on different time scales (e.g., Juarrero (2009), p. 99; Newell et al. 2009, and the references given there). As a rule, the lower the layer, the faster it works (just compare representative physical, chemical, biological and social processes). Therefore, Libet's results can be interpreted in an elegant way: The basic sensorimotor tier quickly sets a behavioural default. Bottom-up, this fixing appears in the conscious mind as a decision, although, in this case, it is just an *a posteriori* rationalization. However, operating on a slower time scale, but being truly in charge, top-level consciousness may readily overrule the lower tier's move (e.g., Donald 2002, Bandura 2008, Baumeister 2010).

Self-organization and order parameters are an elegant way to explain why top-down control is the rule and not the exception. In general, the overlying tiers act as powerful 2nd order constraints influencing the subjacent layers much more than vice versa. If the bottom tier is the sensory realm, and the upper tier the cognitive realm, this corresponds nicely to the well-known view of Kant (1781) that "freedom, in the practical sense, is the independence of the will of coercion by sensuous impulses." If the uppermost tier is the self (see illustration I17), "free will" is an appropriate subjective description for the (partial) "submergence" of the lower layers. To this end, language is an excellent tool since it provides explicit knowledge representation, concise chunks of information that may be combined in a transparent way, yielding resilient lines of argument that may lead to consistent action. It is no coincidence that clear conceptual thinking and understanding, arguing, modelling and checking ideas are so important for us.

Very often, activities are first located on the top tier and subsequently delegated further down. Learning some complex task, e.g., driving a car or playing a musical instrument, starts on the conscious level. One has to understand in great detail what kind of movement of limbs is required, when which movement is appropriate and how the arms and legs interact. Thus the proverb that all beginnings are difficult: they are slow and tedious and take enormous effort. Yet a major part of learning consists in *automatization*. Experienced drivers change gear without (conscious) thinking, and once a pianist has learned a musical piece their fingers know how to move. The saying that some faculty is "ingrained" or has become one's "second nature" captures perfectly what is going on: The skill is deeply rooted within the body, and control by the upper

layer may be restricted to a bare minimum, e.g., a trigger. It is well known that typically (at least) ten years of thorough practice - 10,000 hours of training - are needed in order to learn some demanding activity "by heart". An engineer would say that this time is needed to replace (slow) software with (fast) hardware, i.e., on the basic level, neural networks have to be restructured and programmed in order to master some specialized task.

How strong is the loop within the uppermost layer? Roth (2003) remarks that a vast proportion of neurons in the associative cortex (up to 99%!) communicate with one another. Given this finding, some have concluded that we are constructivists, mainly revolving around ourselves, and building our own world. However, this conclusion is premature. First, in the course of evolution, those who forgot the outside world did not survive. The same result would occur if we could voluntarily influence the output of our sensory devices, i.e., perceive the world as we would like it to be. Second, the top level is thoroughly based on all the other layers below; it is not a "spirit in the sky". Third, the lower layers' input is still important if not decisive, if "informational updates" of the top layer are frequent (e.g., several times a second, say), and if this input influences the internal (circular) processes in the top level sufficiently. That is, in order to have a stable flow of information it would be straightforward if the circular processes in the top level reached an equilibrium without external input. However, if it is mainly the impulses of the lower level(s) that gives them direction, the final result (i.e., the top down arrow on the left hand side in Illustration I15) may depend on the input in a crucial manner. (Although Lady Macbeth is just whispering - not shouting - most of the time, she has a major influence on the overall plot!)

Constructivist ideas, emphasizing the internal processes of the top layer, *underestimate* the influence of the "bottom up" input. This could also be why some cognitive therapies, aiming mainly at the conscious level, are not as effective as one would wish them to be: Talking about depression won't make it go away; however, sports, aiming at the physical and emotional tiers is much more effective.

Of course, if the flow of information between layers breaks down, circular causality is destroyed, leaving the layers unconnected and thus dysfunctional. But pathologies already arise when the physiological dynamic equilibrium between layers shifts. On the one hand, it is typical for many psychosomatic diseases that a lower tier has spontaneous impact on a higher tier. For example, panic attacks strike, that is, all of a sudden a person is overcome by fear, and a major symptom of schizophrenia is uncontrollable sensory impressions, e.g., voices speaking up, or non-existent persons coming into sight. On the other hand, the influence of the upper layer may be too strong, resulting in obsessive compulsive disorders. For example, anorexia nervosa is characterized by an obsession with controlling the amount of food eaten. Cognitive control is way too strong, overruling the sensation of hunger's influence on food consumption. Using the metaphor of the self riding a horse, the first class of pathologies is characterized by a mulish horse that time and

again threatens to unsaddle its overchallenged rider, while the second class may rather be characterised by a reckless rider on an overloaded horse. For more psychopathological examples see Kelso and Tognoli (2009), p. 1112.

D) Nature's *bauplan*

In general, we have described and studied multi-layered (hierarchic), dynamical, self-referential and, to a large extent, also self-organizing information-processing systems, situated in a complex environment (see Freeman (1999) for a similar account). A fully functional mind is a well-orchestrated, multi-modular organization; each and every part having its well-defined place and task, and embedded in a multitude of loops. The overall result may be displayed in a single picture:

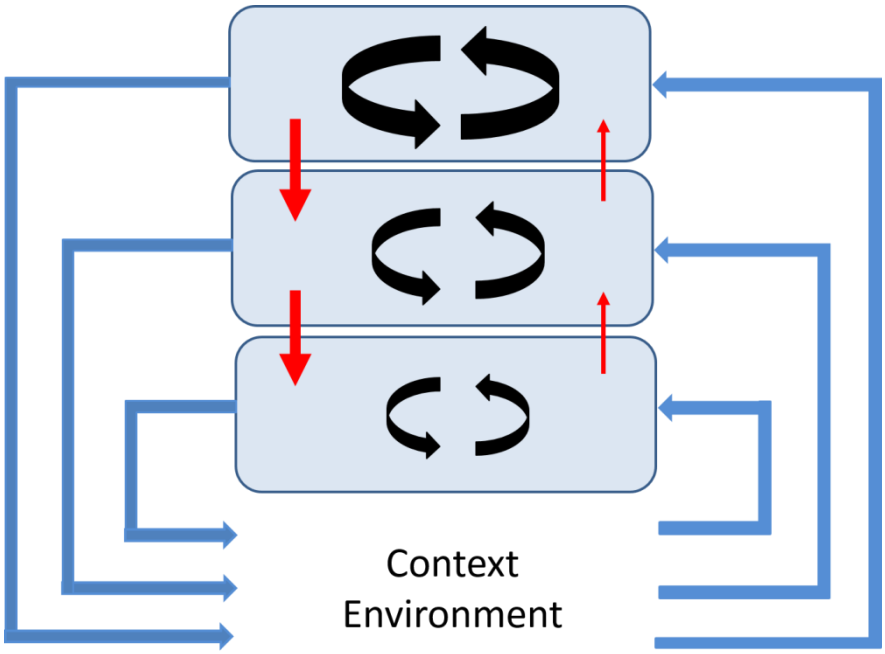


Illustration I18 (A complete multi-layered system with three kinds of feedback loop)

In a nutshell, such a system consists of several layers. The locus of control is towards the top, that is why the internal processes (black arrows) there are more important than those further down. The same holds for the interfaces between the layers (red arrows): The influence "top down" is stronger than the influence "bottom up", therefore the difference between ↓ and ↑. Moreover, there are sensorimotor loops (blue arrows). That is, all layers may cause actions (left hand side of illustration I18). If a certain action changes the environment of the system this change subsequently alters the sensory input, possibly for all tiers (right hand side of I18).

Note that, given embodiment (implying circularity) and modularity (leading to several tiers), the above construction plan is almost inevitable. These basic boundary conditions imply that successful natural and artificial systems have to be constructed in the way displayed in illustration I18. It is also straightforward that such a system has three major modi operandi:

1. Fully aroused (awake), i.e., all three kinds of loop are transmitting large amounts of information. In particular, there is a strong connection between the robot and the external world. For the reasons given in section 1, a human being is ego-conscious in this modus, and illustration I17 applies.

2. Asleep (light sleep), i.e., information flows mainly over the black and red pathways, while sensorimotor loops have been largely shut down. Since the distinction between inside and outside does not exist in this modus, there can be reality-oriented dreaming, at best, but no clear consciousness. Since the mental system is still connected, various brain regions may interact rather substantially. In particular, information obtained while awake could be incorporated into (cross-areal) neuronal structures.

3. Deep sleep, i.e., with only the black arrows being active, information is mainly processed locally, within layers or modules. Like a shop that closes temporarily, this "fragmented" modus allows for major reorganisation and repair, even on an elementary level. Of course, since complex mental functions hinge on connectedness, there is no kind of consciousness in this modus.

E) Differentiation and integration

The dynamics of such a system can be characterized by *metastability.* Kelso and Tognoli (2009), pp. 105-108 explain: "One theory stresses that the brain consists of a vast collection of distinct regions ... The other school of thought looks upon the brain not as a collection of specialized centers, but as a highly integrated organ... metastability is an entirely new conception of brain organization ... Individualist tendencies for the diverse regions of the brain to express themselves coexist with coordinative tendencies to couple and cooperate as a whole. In the metastable brain, local and global processes coexist as a complementary pair, not as conflicting theories. Metastability, by reducing the strong hierarchical coupling between the parts of a complex system while allowing them to retain their individuality, leads to a looser, more secure, more flexible form of function ... No dictator tells the parts what to do. Too much autonomy of the component parts means no chance of coordinating them together. On the other hand, too much independence and the system gets stuck, global flexibility is lost."

Supporting this point of view, Chennu et al. (2014) write: "Theoretical advances in the science of consciousness have proposed that it is concomitant with *balanced* cortical *integration and differentiation*, enabled by efficient networks of information transfer across multiple scales." Thus numerical measures of dynamic complexity in general and for consciousness in particular have been proposed (Seth et al. 2011). Kelso and Tognoli (2009), p. 112, conclude "A delicate balance between integration (coordination between individual areas) and segregation (expression of individual behavior) is achieved in the metastable regime ... In a critical range between complete integration and complete segregation, the most favorable situation for cognition is deemed to occur ... measures of complexity reach a maximum when the balance between segregative and integrative forces is achieved."

Using the idea of differentiation and simultaneous integration opens up another, rather straight, road to self-awareness. In general, starting with a certain structure, new modules may evolve. Typically, they are at first rather primitive but upon elaboration and segregation they obtain a certain "life of their own". However, since the mental edifice is strongly interconnected, they are also readily integrated into the system already in existence. This happens spontaneously and on all levels:

Given a single sense organ, it is well known that the first tier of sensory cells is occupied with restricted tasks (e.g., the direction of objects) and very limited areas (e.g., a certain spot of the visual field or a certain frequency of sound). Moving up the levels of analysis, the information is integrated, e.g., the areas of the visual field covered get larger and larger. Finally, all unimodal information is integrated into a comprehensive map, i.e., the world as we see, or hear, or smell, or feel it. The next "natural" step, of course, is to integrate all modalities into a comprehensive sensory model of the world, i.e., the world as we see, hear, smell *and* feel it.

Within a sufficiently rich perceptual model of the world, there is a prominent token: The self-image, i.e., a comprehensive map, representing the body in the perceptual realm. Combining this map with all the available information about the (inner) states of the body and motoric agency, readily yields a comprehensive body schema. When vocabulary and its accompanying structure evolve, a new module comes into being – language. Soon, within this area, there is a pronounced token for oneself, the word "I", say. With the integration of language into the overall system, it is almost inevitable that the new word is connected with the existing body schema (the integrated representative of the entire body so far). This fusion creates a conceptually sharp and stable distinction between oneself and the rest of the world. In other words, distinct self-awareness, an(other) emergent entity, appears, further triggering the drastic effects already described.

This train of thought underlines that it makes sense to distinguish between the protoself, the core self and the extended self, since each of them is located on a different mental tier. However, looking at the structure displayed in the rightmost column of Table T2, these selves also build on

each other. More precisely, as a result of integration, the protoself is an integral part of the core self, the latter being a crucial part of the extended self. Finally, consolidating all available representations yields a single, comprehensive concept of oneself – one's self, embedded into a larger context (see illustration I17).

On a more abstract level, Juarrero (2009) says: "Dynamical closure always generates a boundary between the new emergent and the background. In the case of autopoietic structures the boundary is self-created by the very dynamics of the system. It can take the form of ... a dynamic phase separation between the emergent structure and the environment, or between the structure and its components."

F) Universal building blocks

Illustration I18 points out that, despite the nontrivial bauplan, there is a single universal building block, used over and over again: it is the information flow through circuits or *feedback loops*, also called "closed-loop causality" and "circular causality". This building block appears in various guises:

1. Sensorimotor loops, connecting the outside world with the internal mental life

2. Circuits providing the information interchange between contiguous layers. In IT-jargon, the contiguous upper and lower layers very much act like a client and a server

3. Loops within the tiers and modules, in particular loops serving as interfaces between modules, and loops mediating parts-whole relationships (e.g., between modules and their sub-structures)

Since evolution "likes" to reuse ("re-cycle") approved building blocks, it is a straightforward conjecture that *all* information processing in natural, self-organizing information systems is heavily based on feedback loops, from the processes within neurons, to small neural nets, neocortical columns, larger modules and networks, cerebral areas, complete functional systems, the integral brain, and – finally -- the whole nervous system.

Moreover, every biological unit, but also every robot, is situated in some context. Thus the very first loop, that is, the elementary sensorimotor loop connecting the "machine" with the outside world, is inevitable. It fits well into our understanding of evolution that this very first loop was re-used, modified (split, differentiated, put to a different use, redirected, strengthened, weakened, dissolved etc.), and gradually extended. Thus creating specialized modules, distinct areas and hierarchic layers, all of them tightly linked, and combined in an overall sound architecture,

"thinking" (more and more sophisticated internal information processing) developed. Finally, well-orchestrated mental edifices with a clear understanding of themselves and their situatedness appeared.

The existing literature places much emphasis on "downward determination" and "circular causality". According to the reasoning in this article, these terms are important, however, they may also easily miss their target. First, the best formal account of causality is based on directed, *acyclical* graphs (Pearl 2009). Second, although it is correct to acknowledge the role of top-down processes (giving the higher layers at least some influence), one should not overlook the fact that each of them is just a part of more important information-processing loops. The same with the idea of a "closed loop". Of course, by definition, every loop is closed, i.e., the end-point of some process coincides with its starting point. However, there is also contextual input and procedural output which may be crucial. In this sense, information processing loops are open, they interact massively with one another and their environment. Third, causation and determination are often contrasted with chance and freedom. Since there is an abundance of reasons and causes and since, traditionally, free will has been associated with non-determinism, one is easily led down the primrose path of fundamental discussion.

This author thinks that dynamical system theory should take centre stage, as its emphasis is on the behaviour of complex systems. Thus it is preoccupied with systems being composed of many particles and being held together by "forces" of all kinds. Moreover, context and constraints play a major role, and one has to consider numerous and diverse factors, be they deterministic or stochastic. The modes of such systems range from straightforward convergence, and (quasi-)deterministic behavior to arbitrary random fluctuations with all kinds of regularity and irregularity in-between (e.g., periodicity, more or less stable attractors, turbulence and chaos). There also seems to be self-organized criticality (Èrdi 2008, Sornette 2009, Fingelkurts and Fingelkurts 2013), in particular, when a certain state of mind - in this view a certain attractor - becomes unstable due to saturation, self-amplification (Haykin 2013, p. 442-443) and resonance, e.g., when the best fitting option supersedes all others. Several authors remark that the brain seems to be working "close to instability points" or "at the edge of chaos" (e.g., Chialvo 2007, Legenstein and Maass 2007), when information throughput and complexity are highest. A thought-provoking application of these ideas to our subject can be found in Andrade (2008) who sees a hierarchy of regimes: Physical Information Systems, Information Gathering and Using systems, and Hierarchical Dynamical Information Systems.

What is crucial is the flow of information. This flow is organized in myriads of feedback loops, all of them working simultaneously but at the same time being heavily (hierarchically) interconnected. In the style of Swift's society of fleas, a loop has smaller loops on which it relies and still larger loops that build on it. In addition, this massively parallel, "Goldilocks-like" - not too tight, not too loose, cf. Juarrero (2009) -, and "metastable" (Kelso and Tognoli 2009,

Fingelkurts and Fingelkurts 2004) processing of information is dynamic: It always changes, never converges or comes to an end. Instead, at any one time, there is some amount of activity which is also variable. However, although the content and the intensity of the internal course of events alter ceaselessly, and, at times, almost unpredictably (e.g., due to new input), the mental stream is kept within certain bounds. In a deep sense, thinking is like (endless) weaving, with elementary mesh loops combining into patches, models and cloth. That's the bottom-up view. However, at the same time there is "downward causation." That is, the whole "loom" (i.e., the entirety of all meshes) and the patterns it produces blaze the trail for subsequent activities on lower tiers.

G) Some maxims

The ultimate challenge consists in building an autonomous machine endowed with a self-extracting multi-layered control system, i.e., to create a mentally developing robot (e.g., Weng 2004, 2007; Cangelosi and Schlesinger 2015). To this end, it seems helpful to ask how nature succeeded in programming its "survival machines" (Dawkins 2006). We have already mentioned her massive parallel approach. Ceaseless as these innumerate processes may be, computation costs time and effort (energy, resources, etc.).

Therefore, a **first maxim** must be to minimize this expense. Haken (2004), p. 17, gives a nice example: "It is often believed that in [the] recognition process an enormous number of details are analysed … The evolutionary process suggests the opposite." More generally, it seems appropriate only to "think" as much as necessary in order to get a desired result. In other words, elegant solutions restrict central information processing to the inevitable minimum and take advantage of the physics of the body as well as the services of the environment whenever possible. More precisely, nature's economical recipe seems to "shift the computational load from the [central] controller to the morphology and physical properties of the embodiment … The controller is challenged to maximally exploit the physical peculiarities of the body in its interaction with the environment." (Der and Martius 2011, pp. 29).

Efficient management of a robot uses its scarce resources optimally, that is, it externalizes burdens whenever possible, maximizing the attainable effect but minimizing internal costs. Paradigm examples can be found in Der and Martius (2011). Crucial ideas are collected in Brooks (1999) who underlines that it is embodiment that provides meaning (semantics), that a successful robot needs extensive front and back ends (i.e., powerful sensory and motoric devices), that very often the world is its own best model, and that intelligence is rather determined by the dynamics of the interaction with the world than by explicit representation and reasoning.

The **second maxim** is to start with simple building blocks and to use them time and again, tailoring them to some specific need. Adaptive neural networks, connected by ubiquitous feedback loops embed the individual in the outside world, but also assemble neurons into small, big and huge units – from neocortical columns to brain hemispheres. Although these units' structures cannot be identical – since they have to cope with different problems - they all work on similar principles, and need to be integrated if necessary. For example, it is well known that pattern formation is almost identical to pattern recognition, and similar to decision-making (see numerous references to Haken throughout this contribution). Visual and auditory perception are "a tale of two sides" (Haykin and Chen 2007). Moreover, temporal binding of brain areas always depends on spontaneous synchronization (Fingelkurts and Fingelkurts 2004).

In a nutshell, there are countless neural networks, myriads of modules, and several layers, all acting in parallel and simultaneously. The formidable task of fine-tuning is mostly solved via hierarchic and dynamic self-regulation, channeling the flow of information. Since timing is crucial, so are "spike trains" (Gerstner and Kistler 2002) and their precise synchronization (Haken 2008). Memory is also organized in a unified way, with content being distributed throughout a net of neurons rather than put in a single "drawer" at a particular location. With the information being laid down in the matrix of neuronal connections, memory is dynamic and self-organizing, with some input evoking a certain dynamic response, typically resulting in a fitting output.

In this view, the basic functional unit is a module, i.e., an array of connected neurons. It is rather obvious that such a functional unit can be programmed in two completely different ways: On the one hand, there is "normal" plasticity. Upon gradually strengthening or weakening the connections within this group of neurons, memory or any other function changes slowly. However, on the other hand, there is also "fast learning", especially during sensitive phases. A plausible mechanism to this end is "massive pruning", i.e., to start with a large number of neurons and links, and subsequently eliminating most of them during the learning process, resulting in hardware that has been customized quickly to a certain context.

These completely different ways of putting a module into operation may explain the enormous differences upon learning a similar task, e.g., between first and second language acquisition. Thanks to the first maxim, i.e., since it is costly to first build a large field of neurons and then destroy most of it, the later process should be the exception in normal (adult) life. However, when the focus is on rapid development, i.e., in children, the second process should be widespread, and explains in part why they need such an enormous amount of energy to build up their mental edifice.

Since learning is tedious (consuming time and energy), one can also expect that nature uses prior information whenever available. That is, pre-structured neuronal networks, ready-made for a

specific task, should be ubiquitous. On such a basis, learning rather resembles grouting and fine-tuning than a major effort which is inevitable when building a structure from scratch. The example of language acquisition demonstrates the enormous difference: Within a short sensitive phase, children learn to master their mother tongue better than adolescents do a second language. Moreover, hardly any amount of training after the sensitive phase will suffice to reach the level of command a child has obtained in passing.

The **third maxim** is to use self-organization wherever possible. For example, instead of teaching a robot many special tricks or having him store one sensory impression after the other explicitly, it seems much more advisable to compress the necessary information to a bare minimum and re-establish the original when necessary. The popular format MP3 does not store a song completely. Rather, it stores and compresses the information relevant to the human ear, ignoring the remainder. It is also not necessary to save an image completely. Rather, it suffices to retain some particular features and fill in the rest upon request, i.e., given certain clues. The human eye is also not a camera taking picture after picture and combining them into a movie. Rather, elementary saccades look for differences and just update those parts of our view that have changed.

Haken (2006), p. 28, summarises: "Quite often it is assumed that the incoming pattern is compared with templates. However, the storage of a template would require quite a large amount of information. Therefore, one might imagine, in the sense of synergetics, that only specific characteristic features are stored in the form of order parameters which may then be called upon to generate a detailed picture. In this sense then, pattern recognition becomes an active process in which new patterns are formed in a self-organized fashion…"

It may be added that every conventional computer program can be understood as a compact recipe to some end. Upon its execution, that is, upon putting it in a certain environment, it is decompressed and creates all the effects it is supposed to produce. Interestingly enough, Turing showed that very few building blocks (in particular loops and bifurcations) suffice to compute anything that is calculable. Notice the deep-rooted similarity: Computer programs, genes, inseminated egg cells - indeed any kind of offspring - are seeds that, if put into an appropriate context, develop rather automatically, they "unfold" there so to speak. However, self-organization goes much further.

First, due to the permanent feedback of the organism and its environment, self-development is strongly adaptive in the sense that the course of "unfolding" is very much guided by local, specific boundary constraints. For example, given the initial competence of language acquisition, every healthy child is able to learn any language perfectly, just depending on the area where it grows up. In the extreme, the context acts like cladding being filled with the evolving structure.

Second, development is automatic and follows general rules: It always starts with crude, rather rudimentary beginnings, e.g. immature neural equipment. Given a reasonable context, however, humble abilities differentiate into sophisticated ones. An appropriate amount of guidance and protection certainly helps, yet most of the construction work has to be delivered by the developing structure. Moreover, depending on the ability to be acquired, there are more or less restricted time slots. Typically, it is much easier to learn a skill earlier in life, when the brain and the body are "made for" the acquisition of new faculties of all kind. Since abilities typically build on each other, there is also a natural order in which skills should be acquired. It is futile to teach mathematical subtleties when the pupils have not yet understood elementary numbers.

Third, despite all the work that is going on, upon gradually extending the system "loop by loop", the whole system remains robust. One could call this "self-organized stability." New modules are established, tested, run in, geared to each other, gradually added to the whole system, and finally used on a regular basis. Again, in a quite self-organized manner, single building blocks form larger structures, until, when the system has matured, all layers are "installed and ready." Trying to design and implement a complete software edifice for a robot thus seems a hopeless endeavor. Instead, nature chose not to build "Rome" in a day, but to have humble beginnings grow and thrive.

Fourth, with complex dynamic systems come all kinds of emergent phenomena. In particular, larger aggregates attain abilities that their components do not have. For example, single neurons have a very limited behavioral repertoire, yet neuronal nets can store information and compute complex functions. The components of a cell are just biochemistry, yet the cell can replicate, i.e., manufacture a copy. Multicellular organisms can differentiate, forming versatile bodies with astonishing features. Such "phase transitions" when "completely new dimensions" are reached, are not the exception, but the rule. They happen quite often and all of a sudden. The popular idea of *self-organized criticality* (SOC) even suggests that evolving systems may have – or attain - the ability to provoke such "tipping points" (see the vast literature inspired by Bak et al. (1987)). Typically, the new properties are almost unpredictable and, at best, explainable with the wisdom of hindsight. Nevertheless, they may have dramatic consequences. The amazing phenomenon of self-awareness fits perfectly well into this global picture.

(*Continued on Part IV*)