Level 7

Cybernetic Universality and *Lives in Phenospheres*

It seems that *living systems* carry the intuitive connotations that characterize cybernetic systems in general. The key question then becomes *what is life.*

The problem with life sciences is that there exists only the one example, the carbon-based DNA life available to us for analysis. The goal here is to extend from *life as we know it* to *life as it could be,* from traditional biological systems to "bio-logical" ones, where the logic follows the relevance-oriented, cybernetically motivated lines of thought. Indeed, one could define *universal life* as *higher-order dynamical balance in the environment,* whatever that environment happens to be. The definition covers, for example, living systems in the chemical environment (lower-level life) and and in social environments (higher levels of life). Because of the very compact structure of cybernetic models, different systems become formally analogous, and when interpreting a system's behaviors in another framework, some fresh intuitions can be available.

Concrete examples are the best way to present the challenges. In this chapter, *infosphere* will be exclusively studied: After all, the *cognitive system* is well understood — or, at least, it has been studied very much. Specially, it will be studied what is the interpretation and essence of the sparse-coded mixture models (chapter 6) in that domain. And perhaps understanding the peculiarities of systems in such a complex environment helps to see the possibilities of evolving life in general: Indeed, when a living system is defined as above, the universe is full of strange life forms — literally, *what you can imagine, it exists.*

7.1 Modeling of cognition

The neocybernetic framework not only allows modeling of the coding of individual patterns, it can perhaps give tools to attack the functioning of the complete brain. There is the intuition backing up us here: The cognitive system simply *has to be* cybernetic — even in various ways (see [30], [53]).

7.1.1 Population of neurons

For simplicity, assume that the brain consists of identical neurons that follow the Hebbian learning rule [37]. It is evident that Hebbian learning exactly follows the same evolutionary learning principle as presented in chapter 3: If the neuronal input and the neuronal activity correlate, the synaptic strength increases. Indeed, the Hebbian neurons are paradigmatic examples of cybernetic agents, the "resources" now being the incoming signals. Employing the idea of looking at the neurons as a population of competing individuals, one can see the neuronal "ecosystem" as a self-regulative entity. No central control is necessary, nor some "operating system", it is all about a distributed agent-based pursuit for activation. This competition becomes is very concrete: It has been observed that there are *nerve growth factors* that control the "wiring" of tissues; here it is the winner neuron(s) only that prosper, and become connected.

But it is the inter-neuronal connections where an especially delicate control is needed. Let us study a scenario. Suppose that there is a pool of more or less occupied neurons available competing for activation. If there is currently very little activation coming from outside to a neuron ($\mathrm{E}\{\bar{x}_i^2\}$ remaining low), the neuron's internal feedbacks make it search for more activation (the coupling factor q_i increasing; compare to the algorithm in Sec. 6.4.1). The "hungriest" winner neuron (or the winners if there is plenty of activation to share) connects itself to the sources of temporary activation, essentially coupling the simultaneously active input signals together[1]. The winner neuron hopefully becomes satisfied and less "hungry", exploiting the resources (signals) thereafter allocated for them. That neuron (or set of neurons) starts representing the corresponding association structure, defining a (subconscious) "concept atom". If such activation patterns are relevant, if they occur sufficiently often so that the corresponding neurons do not starve in the loss of activation again, these memory structures remain valid also in the long run; otherwise the association is volatile, fading gradually away. As atomary concepts are connected to previously activated ones, sequences of concepts emerge. In the long run, the original time structure becomes ripped off: The sequential chains of neurons becomes a parallel group of simultaneously active neurons, competing for more or less the same input resources, and some kind of a *semantic net* emerges. Because of identical correlations-based learning in all neurons, the connections in the net gradually become bidirectional, and an "associative medium" is constructed, being available for yet other (still more elaborate) concept atoms to be connected to the available activity centers in the medium. Lower-level concepts are inputs to higher-level concepts — but as time elapses, structures become cyclic and more or less blurred, the network becoming "panexplanatory".

This all is more or less trivial — the added value, the main contribution of the neocybernetic perspective, comes from the ability of explaining how the above declarative representations change to associative ones, or how the *shift from novice to expert* can be explained. The key functionality is the self-regulation and self-organization property of the Hebbian feedback system: As the Hebbian adaptation takes place, locally and independently in each synapse, the declarative structures become swallowed in the associative medium. As correlating

[1] The overall activity of the network remains constant — when there is no external activation, as in sleep, the system becomes activated by random noise

concepts are appropriately connected together, links of the semantic net become denser and numerically optimized.

The above process of automatization is the key process in the mental system. This all sounds very simple, even a bit simplistic, and indeed this is not the whole story. The mental machinery is not there only for data packing.

7.1.2 Role of semantics

When proceeding from the level of signal processing to processing of information and knowledge, one is facing new challenges, because one needs to address issues that are the most relevant to the human mind: A cognitive model is void, its essence escapes, giving rise to *Chinese room* type arguments [70], if it does not somehow capture the *semantics* of the constructs. One needs to extend from the infosphere, where it was simply data (co)variation that needed to be captured, to "ideasphere". The units of information to be modeled are indeed *knowledge;* mental models should somehow capture "information flows of information".

Cognitive functionalities, like *intelligence,* are emergent phenomena. It is assumed here that intelligence is an illusion that emerges when a large number of simple structures cumulate. For analysis, one needs to be capable of reductionistically decomposing the cascaded system hierarchy into self-contained entities. It is here assumed that the principles remain the same also on the new emergent level, so that the processes can be reduced back to processing of data. Now, assuming that these simple structures are individual cybernetic models for subdomains, how to avoid the infinite recess, concentrating on a single level, truncating the succession of models? In other words: How to assure that the data delivered to a cybernetic system constitutes a "cybernetic closure"? How to fix the grounding of semantics, or make the concrete data contain the necessary "atoms of semantics"?

The concept of semantics needs to be formalized at some level. When processing signals, the relevant information being expressed as (co)variation, one concentrates on *contextual semantics,* where the meaning of the structures is determined in terms of their interconnections, finally reducing back to the system inputs (*naturalistic semantics*). For a cybernetic system, however, this kind of static definition is not enough, one once again needs to extend the studies to dynamic domain to have a grasp of *cybernetic semantics.* It was balances that were the key issue in neocybernetics, and the cybernetic models are models over such equilibria. These balances need to be buried in data, or, the data needs to be made balanced.

Again, it is the dynamic equilibria and tensions that are the basic notion here. In each state there is a tendency to move in some direction; this "flow" is proportional to the unbalanced tensions in that state, and can be utilized to quantify the counteracting forces. Such tensions are also visible in the observation data: State changes, or differences between successive states are proportional to the flow. When such derivatives are included in the data, they represent the additional compensating forces and using them it is possible in that state to reach

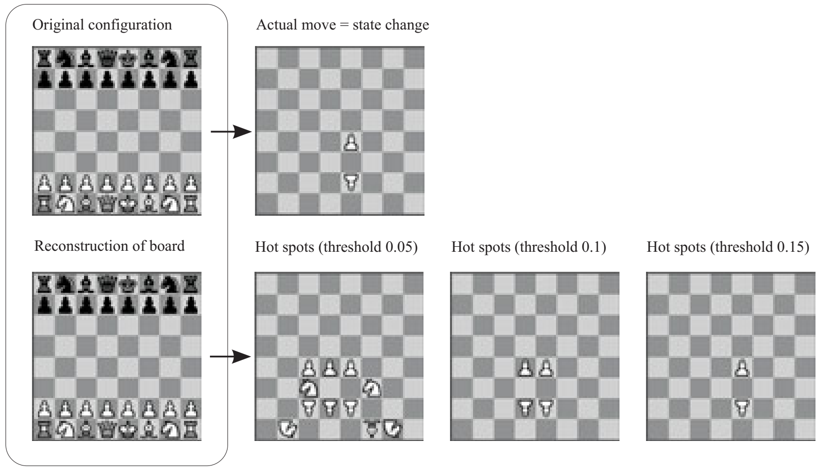

Figure 7.1: Reconstruction of the board and visions of the future in the beginning of the game

the cybernetic balance among data:

$$u'(k) = \left(\frac{u(k)}{\frac{du}{dt}(k)} \right) \approx \left(\frac{u(k)}{u(k+1) - u(k)} \right). \tag{7.1}$$

Such "preprocessing" of observations, emphasis on changes or differences between successive ones, can also be motivated in terms of psychological and neurophysiological studies — constant inputs become saturated, changes in sensors are better detected. From the control point of view (see 7.2), there are also connections: If the variables contain the derivatives in addition to the absolute values an (extension) of (multivariate) PD control can be implemented. Mathematically one could speak of the complete set of variables as spanning a *phase space.* Comparing to mechanical systems, if the original variables are "generalized coordinates", together with the derivatives they determine the system state.

As an example of the relevance of the above discussion study a case where chess configurations are modeled. Chess is the "banana fly" of cognitive science, being a simple domain, but still being far from trivial. There were some 5000 configurations from real games used for modeling[2]. The coding of the configurations was carried out so that for each location on the board (altogether $8 \times 8 = 64$) it was assumed that there are 12 different pieces (at most) that can be located there, and for each of them there was a separate entry in the data vectors. This means that there are altogether $64 \times 12 = 768$ binary entries in the highly redundant data vectors — and when the derivatives were included and u' was defined as in (7.1) the data was $2 \times 768 = 1536$ dimensional. The sparse coding algorithm in Sec. 6.4.1 was applied for the data with $n = 100$, so that 100 *chunks* (as the memory representations are called in cognitive science) were extracted. After convergence typical chess configurations were reconstructed as weighted

[2]I am grateful to Professor Pertti Saariluoma for the data material and for encouraging discussions

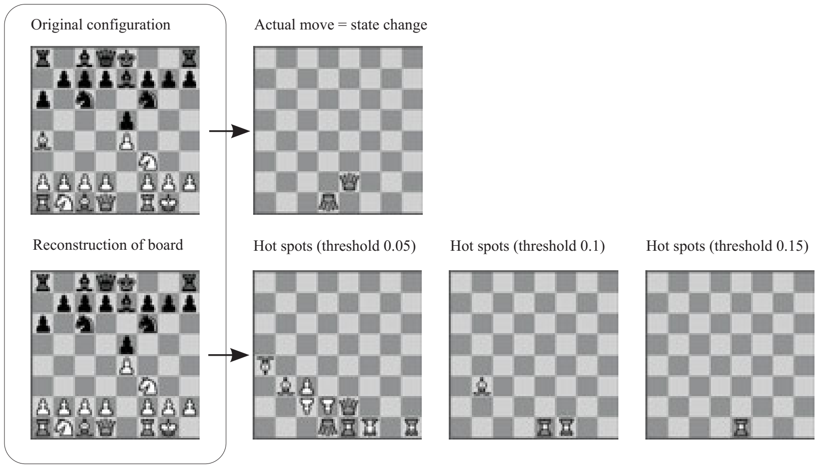

Figure 7.2: Reconstruction of the board and visions of the future at some later phase in the game

sums of the chunks: In Figs. 7.1 and 7.2 the results are presented. Visualization of the high-dimensional data structures is a challenge — in the figures, the modeling results are approximatively illustrated by projecting the numeric representations back onto the discrete-valued board realm. On the leftmost images in the figures, the observed chess piece configurations $u(k)$ are presented: On top, there is the outlook of the original board, and on the bottom, there is the reconstruction when using a storage of only 100 numeric chunks that are appropriately stacked on top of each other. In such a typical case, almost all pieces can be correctly recalled (the vector $\hat{u}(k)$ is thresholded so that only pieces with relevance $\hat{u}_j > 0.5$ are shown). The remaining images illustrate the "flow" of the game, or derivative $\frac{du}{dt}(k)$ in the current state k: Again, on top, there is the observed change in the configuration, and on the bottom, there is the estimate, visualized applying three different threshold levels. The pieces upside down denote vanishing pieces. Note that the reconstruction is purely associative, and no check for validity is here carried out, so that some "ghost spots" also exist. On top of the associations, higher-level reasoning would also be needed to screen the reasonable moves.

It seems that when cybernetic semantics is incorporated in the data, some cognitively relevant functionalities can be emulated: For example, it becomes possible to attack the challenges of *attention.* It turns out that the "hot spots" in Figs. 7.1 and 7.2 are located rather appropriately, and, as it turns out, it is indeed the expert-selected move that has a strong representation. The results remotely remind the mental operationing of a real chess expert: It is known that chess experts only concentrate on the "hot spots" on the board. This kind of attention control has not been satisfactorily explained. Of course, the current experiment only studied very elementary patterns on the board, and to capture phenomena like *functional chunks,* to reach towards really "understanding" the game, one could introduce more complex (cybernetic) preprocessing of the observations [3]:

$$u''(k) = \left(\frac{\bar{u}'(k)}{\bar{x}'(k)} \right). \tag{7.2}$$

It is interesting to note that it has been claimed that some 50000 chunks are needed to satisfactorily represent the chess board [17]. Now the numeric nature of the chunks and inherent optimization of the representations makes it possible to reach a much more compact model for a domain. What is especially interesting is that the errors that the model made were cognitively credible and "expert-like".

7.1.3 Epistemology of constructs

In today's artificial intelligence (AI) paradigms (like in *semantic webs* and earlier in expert systems), it seems that one is interested in *ontologies.* However, the essence of knowledge is not in the objects but it is in the ways of conceptualizing and representing them. What kind of *epistemologies* are dictated by the underlying "wetware"? Or, more appropriately: What kind of structures are dictated by the cybernetic machinery and data distributions? In *Whorf-Sapir theory* it is observed that concepts are the basis of cognitive phenomena; now the emphasis is on the structures beyond the concepts.

First, the mathematical structures can be compared to cognitivistic models. Perceptions are lower-level observations that are filtered through the mental model. In concrete terms, $\bar{x}_i$ determines the relevance of the concept (category/attribute) number i when perceiving the input. As seen in another perspective, the sparse coded momentary weights $\bar{x}_i$ stand for the cognitivistic notion of *short-term memory,* containing "indices" to *long-term memory* constructs. These LTM constructs are the profiles ϕ_i expressing the elementary patterns of exhaustion of available activation. Sparsity is manifested as *STM capacity.* This scheme is completely distributed and locally controlled; the computer paradigm with its centralized registers, memory units, and data transfer among them, can be abandoned in this framework. The cognitivistic emphasis on *constraints* is well in line with the cybernetic assumptions: Without limitations to allocated capacities, there would be no need for optimization, and there would be no need for emergence of abstracted models.

As it is assumed that it is essentially the same Hebbian perceptrons that implement all the functionalities, there is the common neural basis of the cognitive constructs, dictating their structure. The "conceptual spaces" (see [31]) are not based on clusters in the data space but on optimized axes of degrees of freedom determined by the linear sparse-coded features. Because of this uniformity, it must be so that for example *categories* and their *attributes* have essentially the same kind of structure, each determining the other: The resulting epistemology of categorization differs from traditional views (see [66]). Categories being combinations of attributes (features), and the attributes are each other's attributes,

[3]For example, a variable where one would have a weighted sum of all own pieces, minus weighted sum of all opponent's pieces, would make it possible to include a gross evaluation of who is leading the game; tension towards maximum of this variable would directly incorporate the "will to win"

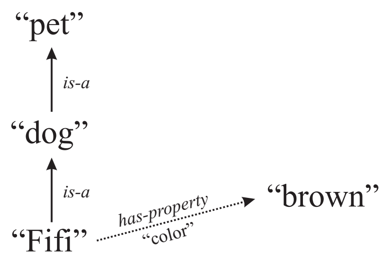

Figure 7.3: How different concepts determine each other: *Semantic net*

determining their contents. Concepts are models abstracted upon examples, so that observations can be explained as economically as possible when employing them; these “concepts” are subsymbolic, but they change to symbolic if their relevances exceed the threshold. Still, all of these structures are numeric rather than symbolic, “fuzzy” rather than distinct, all processing taking place on the numeric level. From the theoretical point of view, it is nice that such collapsing of class structures makes the paradoxes of the Russell type impossible — there are “sets of sets”. As seen from another modeling point of view, it turns out that the “is-a” hierarchies and “has-property” structures become unified. The uniformity and uniqueness of mental structures extends to all levels and conceptual constructs: Also subclasses, and, specially, instances of classes, are similarly represented as interconnected degrees of freedom (see Figs. 7.3 and 7.4):

> A dog is a subclass of a pet, and Fifi is a subclass of a dog — but, simultaneously, a dog is a part of the contents of a pet, and Fifi is part of dog. Inheritance is not hierarchic but becomes a network: Examples of a dog determine what brown color is like, and the concept of brown partly define what dogs are like. Speaking of dogs activates associations to pets, and *vice versa.*

This means that the framework of *fuzzy subsets* offers an appropriate framework for mental constructs — subclasses belong to superclasses, but also *superclasses belong to subclasses.* The subclasses characterize the properties of the superclass to some extent. Fuzziness seems to be an appropriate word to characterize categories, distinct categories are just our way of explicating the world. How colors are see, for example, is dependent of the culture: In some cultures specific concepts do not have relevance. This fuzziness applies also to other cybernetic systems outside the cognitive one. As Theodosius Dobzhansky has observed, “the problem of what is a *species* is among the most acute in biology”. Concepts are just attractors in the surrounding infosphere, or they are not.

The model of the cognitive structures is comprehensive, also including *feelings,* etc. Feelings also have their contents, their semantics being defined in terms of connections to prior experiences — and the contents of other experiences are flavored by the feeling attributes. The difference with feelings is that they seem to be more physical and “deeper” than concepts in general, being bound to the chemical realm: Typically a part of their contents is related to levels of adrenalin, etc. The key point in cybernetic models is that all information is used and all

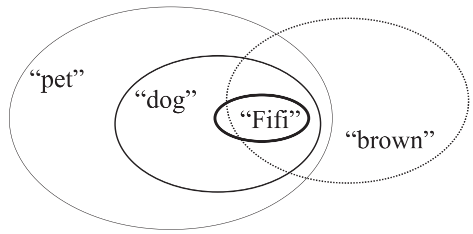

Figure 7.4: How different concepts determine each other: *Fuzzy subsets*

correlations are modeled. When signals are not existing purely in the infosphere but also, for example, in the chemosphere, a tight web of connections to the environment is constructed, constituting a grounding of "self". If associations become reconnected, the contents of the feelings can also change — *neuro-linguistic programming (NLP)* can truly change the way we see the world.

Concepts are names of categories; they are statistically relevant constructs abstracted over individual observations, dependency structures that become addressed, attractor structures that have sustained the tensions in info/ideasphere. The traditional dilemma – the gap between symbolic and numeric representations – is solved because it is manipulation of numbers that makes distinct structures (symbols) emerge from data: Symbols are attractors of the dynamic processes that carry out the data processing. To "bootstrap" an appropriate concept structure, a delicate iterative process is needed. Explicit programming of the concepts is possible, declaratively defining the connections to other concepts, but mere structures with no relevance fade away. There need to exist the structures to instantiate the dynamic processes, but according to the principles of *constructivism,* the structures need to be relevant to flourish. As the poet says: "you can only teach what already exists in the in the dawn of the student's understanding". Without guidance, if the concept formation is completely left to the student (as is the tendency in today's pedagogics), the emergent structures become more or less random, as the syntactic categories cannot uniquely be determined based on the examples alone.

Above, the data samples are identified with "observations" or "sensations", and the results are (artificial) "perceptions" (vectors u and x, respectively), etc. Such direct interpretations of data structures as constructs in cognitive science are rather bold — but in the cybernetic sense they are accurate enough, being *relevant attractors* carrying the correct intuitive connotations, details being ripped off. In the similar manner, there also exist more ambitious consequences that seem appropriate.

7.1.4 On expertise and consciousness

There are many intuitions that are offered by the neocybernetic approach. For example, one can claim that *expertise* in a specific domain is based on appropriate features or chunks existing in the conceptual space. An expert matches the observations against his mental view, thus compressing the data into domain-oriented representations. Applying this low-dimensional representation, missing variables are "filled in" as the known variables are matched against the model, and this way, "associative inference" is implemented (see Fig. 7.5). One

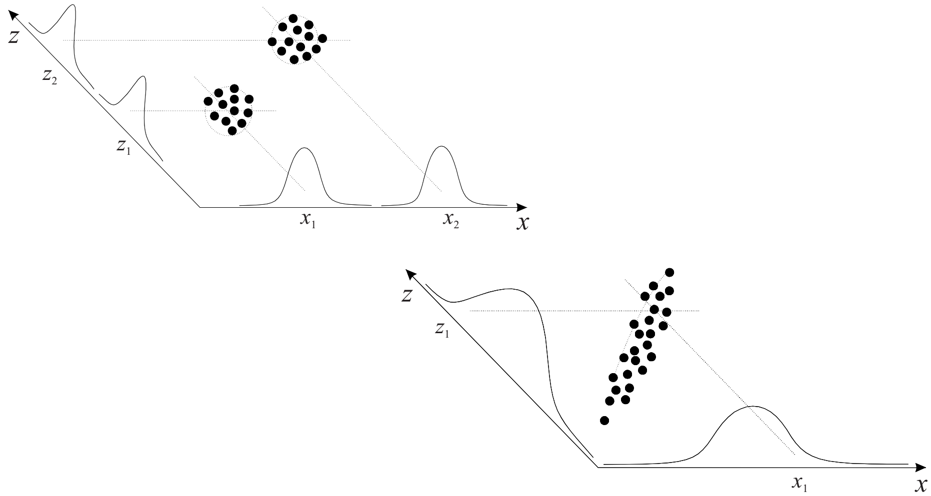

Figure 7.5: Traditional view of expertise (on the top left) makes it possible to implement rules of the form `IF` $x = x_i$ `THEN` $z = z_i$, etc., whereas when distributions are employed, inference becomes subtler, being an associative (maximum likelihood) pattern matching process against the existing knowledge model

of the traditional challenges in artificial intelligence – the "frame problem", not understanding the context — vanishes because of high dimension of data: Even the less important information is carried along in the data structures, the high-dimensional knowledge representations being never isolated from their surroundings. The distribution-oriented view of expertise with continuous fine structure allows subtle, non-binary reasoning, and also merciful degradation of mental capacity as a function of scarcity of resources is manifested.

According to the domain-area experts, or members of Mensa (!), intelligence is the ability of recognizing similarities between patterns — this is what the intelligence tests actually measure. And patterns can be seen as correlation structures. Extracting and modeling of such correlation structures is very much in line with what the cybernetic machinery does. Finding a connection between far-apart correlating units can be said to be an *idea* (or *innovation*), and the general ability of finding such new couplings can be called *creativity.* And when exploiting the Eastern wisdom: In Buddhism, *awakening* is a comprehensive experience, a moment of intuitive, associative understanding.

No explicit determination of the "mental view" is possible — this is due to the limited bandwidth of input channels. Activation of appropriate concepts has to be carried out through a sequential process, sequences activating marginal distributions, gradually spanning "virtual data" in the environment[4]. Similarly, also the output channels are band-limited. Coordinated decoding of associative representations is needed for all communication — not only among people, but also among mental substructures, that is, when *thinking* takes place, when in-

[4]It seems that the transfer of complex information always has to be implemented in a sequential form — for example, when looking at a scene, the saccadic eye movements change the single image into a sequence of subimages that are thereafter reconstructed in the mind

formation is transferred between subsystems that previously have perhaps not been connected. Higher-level tasks, like planning or explicit inference are based on coordinated processing of sequences. Consequently, it is not enough to explain the processes from declarative to associative, or coding of information — also the inverse direction, or decoding associative representations to sequential ones, should somehow be explained by the mental model. This decoding is not so natural process as the coding seems to be: For example a human expert cannot typically explicate his/her knowledge. This would mean losslessly projecting the very high-dimensional virtual distribution onto a set of one-dimensional sequences, natural sentences or formal rules. Such "sequential codes", or languages, will be elaborated on later; perhaps understanding the relevance of languages when trying to understand "living" systems in other domains, too, is the main message here.

There also exist more vague concepts, like that of *consciousness,* can be addressed in the neocybernetic framework. There are many contradicting intuitions of how consciousness should be defined — the heated controversies being, of course, caused by the fact that consciousness is the essence of our specialty among animals. The views vary from the highest (consciousness is the culmination of intelligence) to the lowest level (consciousness is ability to feel something like pain), or even below that (consciousness can only be explained in terms of quantum effects).

Awareness of *self,* or "knowing that one knows that one knows", is assumedly a holistic, emergent phenomenon that cannot be reduced. However, in the adopted framework this structure of infinite recess can again be collapsed. In the neocybernetic spirit, it can be assumed that the mental machinery constructs a more or less sophisticated model of the environment; when this model becomes complex enough, the "self" becomes a relevant entity in the model that successfully helps in structuring the observations and behaviors in the environment. Indeed, when there is a *model of one's own model,* a system can be said to be conscious. This would mean that animals have consciousness in varying degrees — but also non-biological cybernetic systems would be conscious to some extent. On the other hand, a small child not distinguishing itself from its mother is *not yet* conscious — but the "brain prosthesis" can truly capture the mind.

The "cybernetic grounding", or the concretization of the intuitions concerning mental processes as being based on infinite recess, also solves many problems about the *hermeneutic circles.* For example, the traditional definition of *knowledge* is that knowledge is something like *motivated, true belief.* Defining one term then means first defining *three* terms — these terms being, after all, dependent of the concept of knowledge. In the neocybernetic framework the chains of associations *converge* to a balance of referential tensions as they are implemented as stable dynamic processes, the "fuzzy ostensions" being defined by the elements in matrix A. The deepest concepts, too, become matters of scientific study as instead of "truth" the essential thing is relevance: Do there exist appropriate attractors in the ideasphere. Counterintuitively — making the truth relativistic it becomes universal. Similarly, many other age-old philosophical dilemmas can be given cybernetically concrete interpretations.

7.1.5 Theories of mind

As an example of a wide variety of deep discussions concerning cognitive phenomena that are related to cybernetic considerations, study the philosophy of Immanuel Kant (1724–1804) here. Kant was the first to observe that even though we have our subjective mental worlds, there is something objective: Even though we experience the *a priori* existing *noumena* in different ways because we have different senses, we all share the *same predetermined machinery* that processes the observations. Thus, there is possibility of objectivity among people (see chapter 10). Without saying it in modern terms, Kant is actually speaking of models and people sharing the same modeling principles, solving (to some extent) the problem of what is the relation between the external world and internal mind, and how an experience is possible in the first place. He was the most significant cognitive theorist long before his ideas of were coined in psychology — and he indeed was a pioneer of scientific study in this field, criticizing the use of mere pure reason.

One of the basic principles about the human perception machinery is — according to Kant — its capability of constructing *causal structures* among observations. The background here is, of course, the fact that our mental constructs invariably seem to have such a functional structure between causes and effects. This observation has successfully been exploited for modeling (for example, see [62]). However, as observed already by David Hume, one cannot ever see causalities in data, only *correlations,* that is, one cannot without external help detect cause/effect relationships, only simultaneity of phenomena. This seems to be an eternal dilemma when trying to explain the human brain: There has to exist some guiding hand constructing the causalities appropriately, and an external *mind* is needed?

It can be claimed that the neocybernetic model offers a solution to this causality dilemma (compare to chapter 3). Because it is only one's own actions Δu, as induced by the environment, that are being coded in $\bar{x}$, one implicitly knows the structure among causes and effects — there is no paradox any more here. True causality structures are indeed built deep in the Hebbian feedback adaptation strategy: Only models are constructed that are tested in the environment through feedback. The process of true "understanding" is a two-directional process — to truly grasp something, you need to have your "hands on" it, seeing the reaction of the world to your action, as observed also by today's pedagogists[5]. When looking at the cybernetic model (and now one needs to study the "stupid agent"!), the matrix A is not actually any more a correlation matrix but a "causation matrix"; the machinery constructs a pancausal model out from noncausal observations. The information flow from the environment to the system has always been seen as important, but now it is the inverse, or the feedback flow that plays an equally important role: Otherwise there is no emergence of order, and, specially, there will be no causal structures — and this "probing", testing for causality, is built deep in all levels of the structures in cybernetic systems. Note that the causality as seen here is not "trivial" succession on the time axis;

[5]The Finnish words for "to understand" and that for "concept" (or "käsittää" and "käsite", respectively) literally have their origin in the words "hand" and "to process by hand" (or "käsi" and "käsitellä"). Surprisingly, it seems that different cultures have "grasped" or "handled" such deep concepts in similar ways: In German, it is "begreifen", etc.

the variables Δu and $\bar{x}$ find their values simultaneously.

The central role of the self-centered causal models is reflected on the highest levels of consciousness. The sense of *free will* is one's interpretation of what is happening to him/her in the turmoil of the world. Human's mind is built in such a way that when one's intentions match with what truly takes place in the world, one feels like being the subject rather than the object there.

Kant also discusses transcendental arguments concerning the world outside: What kind of properties in the environment are necessary to make construction of the mental model (as he sees it) possible. Even though such discussions are very deep and somewhat obscure, there are simple ideas underneath that still hold; these ideas are contrasted here to the structure of the cybernetic model and the environment. Kant concludes that there are essentially two key properties of the world that are needed:

- **Space.** The observations need to have spatial structure to become manifested as something else than chaos. This ability to distinguish between variables is *implemented through the basic structure of the cybernetic model:* It is assumed that the variables are localized in the vectors, and within the vectors each variable has a distinct role.
- **Time.** Human-like cognitive phenomena are fundamentally based on temporal structures. In the cybernetic models, the time axes have mainly been ignored this far, and in what follows, such extensions are *implemented through the properties of the environment.*

The above starting points nicely draw the borderlines — what kind of models one reasonably can construct and what to ignore. Concerning the spatial structure, there are the basic wiring between the senses and the brains, signals determining the basic dimensions of the space that exists in the brain; beyond that, the assumption of *tabula rasa* can be employed. There are no innate ideas or predetermined faculties in the brain, and the universal modeling ideas should be applicable. But Kant's intuition is deep: Taking the spatial structure only into account is not enough. When attacking the temporal structures, however, the simplicity objective has to be relaxed. In what follows, the time-domain complexities of the real world are hidden outside the system — it suffices to study what it takes if the system is to optimally implement a cybernetic structure in such an environment.

7.2 Manipulating the environment

What is the reason for cognitive systems to emerge in the first place? Nature has not built the mental machinery to think of philosophies.

There is a consistent continuum from basic neurons to the human brain — from the simplest structure to the most complex ones, the objective is always to *change the environment.* This far in the cybernetic studies the system has adapted to match the properties of the environment, but now its role is changed from a silent object into an active subject. Only reacting to the environment, simply trying to survive is not yet what one would think *life* is; intuitively, there must be more *passion* and *free will* involved. To be capable of manipulating the environment in more sophisticated ways, more complicated control structures than what have been studied this far are needed. The following discussion is to be seen only as a demonstration of challenges the cognitive system is facing.

7.2.1 About artificial intelligence

When trying to understand intelligence in wider perspectives, one is entering the zone of (even more) speculative studies. Rather than doing analysis of the environment one tries to make synthesis towards a somehow modified environment. When trying to understand intelligence in general, and when trying to synthesize it, the lessons learned in the field of *artificial intelligence* are invaluable. Indeed, the goals of artificial intelligence are getting nearer to those of cybernetics — sometimes the letters AI are interpreted as *agent intelligence* or *ambient intelligence.*

AI research is a marvelous example of a cybernetic domain where memes compete violently. Cognition, and specially intelligence, are sensitive areas – it is something that is seen as something that is human's own. There are many arguments and counterarguments, the tensions evidently not finding a generally agreed balance. For example, the extreme pessimists claim that *human mind cannot study its own functioning;* on the other hand, the extreme optimists claim that *after twenty years computers are so fast that they beat the human.* Perhaps one should already be capable of outperforming a housefly, then? The periods of enthusiasm and disappointments have alternated, and the whole field has had its collapses and rebirths. It is good to recognize the memes from the past.

There were many starting points for AI back in mid-1900 — one cornerstone was Norbert Wiener with his Cybernetics, and another influential figure was Alan Turing. Indeed, it was Turing that defined the AI paradigm and its objectives: He coined the goal of AI research in his (modified) imitation experiment — a computer is intelligent if it can mimic human [80]. But is it enough that behavior only looks intelligent? This is still today the mainstream approach, but the resulting applications are examples of the "shallow view" of AI, where the intuitive feeling of intelligence seems to escape.

Another contribution of Turing (and other pioneers) was the introduction of the computer metaphor in AI: After showing that the "Turing machine" can implement any computable function, it was easy to assume that also mental functions can be emulated by computer-like structures. Indeed, the standard models for explaining cognition, like the Anderson's ACTR, are still based on memory registers and separate compartments for functionalities [1]. However, the computer metaphor with centralized elements necessarily fails the reality check; there is no explicit transfer of information and no separate localizable memory structures in the brain, but it is all an integrated whole.

But, indeed, Alan Turing was the first to admit that there is more to a zebra than the stripes.

The original approaches to AI seem to be having their reincarnation today: The modern "Brooksian robotics", for example, goes back to very basics of action and reaction structures [11]. In the same spirit, the successes of connectionism have also brought the emphasis from high-level symbolic — cognitivistic — approaches back to low-level data processing with inner structures of no *representation.* Turing's "black box" approach to intelligence has its roots in behaviorism; however, today cognitivism or constructionism are mainstream cognitive science. Should not AI follow here — back towards symbols? Perhaps a synthesis is possible, perhaps it is possible to make the developments a spiral rather than a recurring cycle? The claim here is that the cybernetic framework is the key towards this synthesis.

It seems that the original intuitions about intelligence due to Wiener are still valid: The basic function of the mental machinery is to implement control of environment. But rather than implementing behavioristic control, one can implement more sophisticated model-based controls applying the neocybernetic models with internal representations. When an integral connection with the environment is implemented, "deep AI" can be reached. This connection is not only embedded AI in the traditional sense, but "cybernetic AI".

Implicit control is the basic property of a neocybernetic system. However, now the control view needs to be extended from the side-effect towards the basic functionality. It turns out that some qualitatively new structures need to be introduced, and a certain level of sophistication is needed to support and adapt those structures.

7.2.2 Reflexes and beyond

The assumption here is that when a system reacts appropriately to the environment, illusion of intelligence emerges. In its simplest form, such reactions can be seen as *reflexes,* atomary manifestations of intelligence, representing reasonable behavior — facilitating survival — with no brains truly involved. But there is a continuum towards more convincing functionalities: For example, study the behavior of a *cat* — when it sees something move in its field of vision, it turns its head towards the movement and attacks. In lower animals, like in frogs, such behaviors are still more prominent: Movements in its visual field activate the reflexes. Automated sensor/motor loops can be seen as extensions of simple reflexes, being learned rather than hard-wired, but still by-passing higher mental faculties. As seen from outside, such more or less automated reaction already gives an impression of "real-life intelligence". And this intelligence is reached by a simple cybernetic feedback structure as shown here.

Here, *artificial reflexes,* learned but sub-conscious, are studied, and for that purpose, the originally static model framework is extended to dynamic cases.

Earlier the cybernetic system was seen as a mirror of the environment, environment being mapped onto the system state and from there instantaneously back to the environment, the time axis being compressed into a singularity. Now it is the controller system that implements *current state as a mirror between the past and the future,* and, what is more, an adapted control also should somehow implement *balance between the past and the future.* The cybernetic key principles are still applicable: The goal of the system is to eliminate variation in the environment. When variations in the environment are interpreted as threats, low-level intelligence already has immediate application: Getting away from threats can be seen as control towards zero activation in the local environment. Based on such environmental challenges, emergence of higher and higher levels of mental functions and reasoning skills is evolutionarily comprehensible: Facing a combination of stimuli, should one fight or escape?

The subjective variables are bound to the system controlling the environment — this means that the same goal, or changing of the observed environment, can be reached in different ways: Either by explicitly altering the environment and its variables, or through reaching another viewpoint, meaning that the system *moves* with respect to the fixed environment. The same solutions apply in both cases and the actual mechanisms of how the variables change need not be known by the local controller.

Estimate the future, and when this future is known — *eliminate it,* bringing the future to zero state. As compared to traditional control engineering disciplines, this resembles the *dead-beat strategy.* This kind of control has its weaknesses, including stability and robustness problems in challenging environments, and more complicated control schemes could be studied here, too — however, the dead-beat scheme is taken as the starting point. Still, there exist many ways to implement the cybernetic control depending of the special properties of the environment to be controlled; some alternatives are studied in what follows.

First, take a very simple case that is a direct extension of the original idea of "static control": Assume that if no action is taken, there is no change in the state of the world, so that the future equals the past. This assumption is well applicable in steady environments where change in the variables takes place only through movements of the system. However, responses to one's own actions need to be identified to implement smart controls. To make this simpler, assume distinct, distinguishable excitations, and assume low level of coupling (small q) so that complex dynamics in the environment can be ignored; further, assume that all variables can be affected, that is, with strong enough control, all variables can be zeroed — otherwise there can emerge stability problems in the controller adaptation. To avoid incorrect adaptation, assume that the initialization of ϕ is implemented appropriately. If all these assumptions are fulfilled, only a minor extension to the cybernetic basic model needs to be introduced, namely, *delayed adaptation:* When the control signal $\bar{\bar{c}}(k)$ is there, the matrix $\phi_{\mathrm{c}}^T = q\mathrm{E}\{\delta\bar{\bar{x}}(k+1)\bar{\bar{c}}^T(k)\}$ is updated only after the results of the applied control are visible (see Fig. 7.6). Time indices are used here to synchronize the data, denoting the latest piece of information being employed; the "double bars" are used here because the observations $\bar{x}$ are the inputs into the controller layer; these "single bar" signals are to further find their balance against the new layer latent variables, or control signals $\bar{\bar{c}}$. The model can only be updated afterwards — but it can be applied online because it is the current information only that is employed in control; in addition, the inversion of the control signal because the negativity of the feedback has to be explicitly done, so that the actual control is $-\phi_{\mathrm{c}}\bar{\bar{x}}(k)$.

The right-hand side in the figure represents the process of model construction, where the dependencies between the control signal and the resulting observations are recorded, and the left-hand side represents model usage, or on-line construction of the control signals. The information flows from the "prior input" $\bar{x}(k)$ to the "latent variable" $\bar{c}(k)$, and from there back to the "posterior

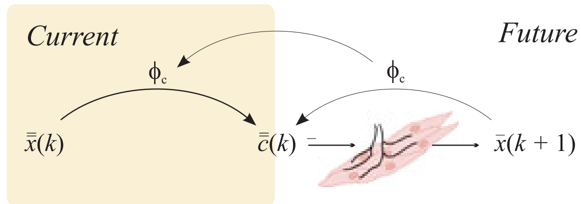

Figure 7.6: Control as a (antisymmetric) mirror between the past and the future

input" $\bar{x}(k+1)$, now assuming that the control signal and the two inputs are in balance — the former balance being implemented by the model machinery, but the latter balance being hopefully provided by the environment. In a stationary environment perceptions in the future are assumed statistically equivalent the perceptions in the past.

The intuitive idea of this control is to make the change in the state *inverse* of the current state, meaning that, when the new state is reconstructed as a sum of the old state and the change, the outcome will be zero state — meaning successful control. The past and the future perceptions are "folded" on top of each other. The procedure can be summarized as follows:

1. Observe the environment $u(k)$ and find the compressed perception $\bar{x}(k)$ corresponding to $\bar{u}(k)$.

2. Using the perception $\bar{x}(k)$ as input, find the balance $\bar{\bar{c}}(k) = \phi_{\mathrm{c}}\bar{\bar{x}}(k)$, and apply the control $-\bar{\bar{c}}(k)$.

3. Update the model by the cross-correlation between $\bar{\bar{c}}(k)$ and $\bar{x}(k+1)$, let $k \to k+1$, and go back to step 1.

In principle, the above scheme defines a framework for mastering independent, co-adapting motor neurons, so that many uncoordinated muscles can do individual "agent control": The coordination emerges as the reactions in the environment are observed. Because of the "humble" nature of adaptations, redundant control structures can be implemented, so that the dimension of c is higher than that of x.

As compared to the standard neocybernetics discussions, some new thinking is needed here: The ultimate homogeneity cannot any more be reached. Structurally, it is necessary to integrate also output (or control) in the models, mere input data processing is no more enough. Above, it turns out that this can be reached easily: An appropriately constructed model for input simultaneously implements an optimized model for output (control signal) construction. Applying this trick, simple structures only are needed as explicit model simulations can be avoided.

Technically, there are extra structures needed to capture the time-domain structure between data. When it is no more one static pattern at a time but a discrete-time succession of samples, some sampling mechanism is needed. And further: Many of the above shortcomings that plagued the presented scheme can be fixed when even more sophisticated structures are employed. Specially, above the details of dynamics were ignored; to proceed, a simple model of the environment is no more enough — a wider view of the world needs to be supplied to the controller. One has to be capable of *simulation,* or estimation of the future in a changing environment, before being capable of eliminating the expected future deviations. The key point here is that there is a continuum from simple to complex behaviors, all new innovations making the control system in some respect better, thus defining a more or less smooth evolutionary path towards extensions.

7.2.3 Extending the mind's eye

In the earlier chapters it has turned out to that a good strategy to inspire new functionalities in the model structures is to take into account the nonidealities there necessarily are. In the similar manner, when studying the extensions that are needed when aiming towards extensions to the cybernetic controls, nonideality of the world being controlled have to be considered. These nonidealities are related to the time-domain structure of real-life phenomena: There is dynamics, being manifested as *inertia,* and *explicit delays* being manifested as latent times between the action and the corresponding reaction. The constraints of the real world become acute and no easy tricks are available: The past cannot be affected any more, and the future is not available yet. To manipulate the world in reasonable ways, to change from an object to a subject, the system has to be prepared to tackle with such properties of the surrounding world.

What is more, the world is characterized by a diversity of variables. There is a multitude of alternatives when analyzing the dependencies among observation entities: Certain variables can have causal structure, or they can be independent of each other. Even if there is correlation among variables, there is no certainty about their mutual causalities — but when implementing control, it is strict causalities only that can be utilized. When deriving the cybernetic models (chapter 3) all variables were assumed to be equally manipulable — indeed, this is what the assumption of "pancausality" is about; this assumption was applied also when implementing the control strategy in the above section. In the real wide world outside the restricted domains, the idealizations do no more apply. The adopted learning principle — increasing the connection strength if there is correlation between the input and the internal state — results in ever-increasing signals, ending in explosion, if the feedbacks from the state cannot bring the input signal down. Of course, these difficulties only become more severe when the dynamic nature of the environment is taken into account: The information about the control effects comes only after a delay, and mere "frustration" of the controller can also result in instability. There are no external critics to moderate the adaptation processes, just as there is no *a priori* knowledge about the causal structure available.

The complex causal structures are the main theoretical problem when striving towards more active controls. Not all dependencies contributing in the outside world can be detected. — But, indeed, the complete structure of causalities is not needed. Note that the observed correlations can be utilized for prediction even though the underlying mechanisms are hidden. There are different routes to the same end result, and it suffices to identify the mechanisms of one's own actions to the future. This way, first mapping the observed current state to the future (applying observed correlations), and from there back to one's control actions (inverting the observed causalities), the general "model predictive

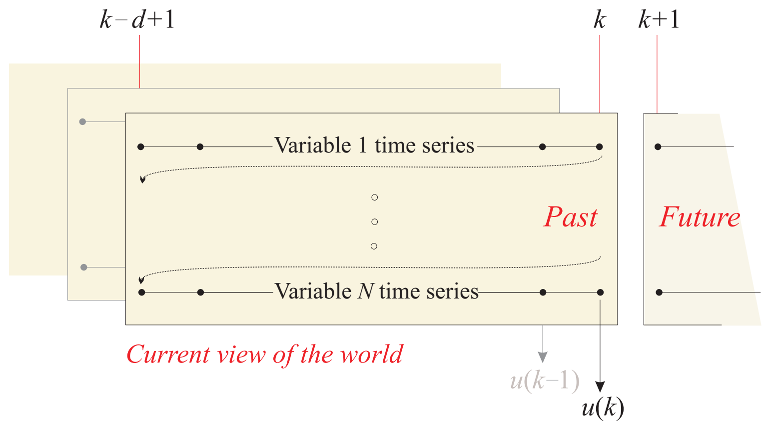

Figure 7.7: How the sensory registers can capture the state of the world

control" can be implemented, when seen in the control engineering perspective. To implement such a scheme, explicit *prediction* or *anticipation* of the future is necessary.

How to represent the past and the future? It turns out that both the past and the future, even though containing infinite amounts of information, can be collapsed into a singularity, and thus can be coded efficiently: According to system theory, an appropriately selected state vector can code the past of a finite-dimensional dynamic system; and if it is assumed that the controls are successful, there is only a short sequence of transients in the future before they are eliminated.

To tackle with the time-domain peculiarities of the world, and, specially, to implement the necessary structures to support prediction of the future, one can employ the concept of a *mental image.* In Fig. 7.7, a simple possibility is presented that can capture the state of the world containing linear dynamics of (at most) order $d-1$. The time series of relevant variables up to current time are assumed to be stored as a high-dimensional vector structure (vector length here $m = Nd$). The sampling interval is assumed to be selected appropriately to capture the natural dynamics. When PCA-like data compression is carried out, the degrees of freedom can be captured — in this case this means that the dynamics of the signals can be losslessly coded. The time series representation makes it possible to express discrete derivatives, so that there is no need to include the derivatives separately among the data (see Section 7.1.2). The control can be based on such mental images: In prediction, the current image $u(k)$ is mapped onto the future image $u(k+d)$.

The representation of the world becomes high-dimensional, and the control strategies need to be robust against redundancy and irrelevant information. But if such robustness can be reached, natural-looking functionalities can be reconstructed: for example, finding correlation patterns among seemingly unrelated observations makes it possible to simulate *conditioned reflexes.*

Getting back from the assumption of extreme homogeneity to tailored structures means that also the assumptions concerning separate structures for sensory memory, different kinds of registers and buffers, etc., become necessary again. The simplicity objective applied in modeling must not override the facts. And, indeed, it has been observed that there are various specialized functional structures in the brain. Processing of dynamic phenomena seems to be a central part of brain functions: For example, the cerebellum is known to be related to processing of low-level dynamics. Model-based simulation, or reconstruction of the future, truly seems to be characteristic to brains (compare to “mirror neurons”, etc.). And the modern brain imaging techniques have revealed that when perceiving dynamic phenomena, there exist brain waves at certain frequencies; it is tempting to assume that such cyclic operation is related to the mind’s internal discretization machinery that changes the continuous flow of sensations into a discrete-time sequence of perceptions.

7.2.4 Implementing more sophisticated controls

When studying the possible control-motivated extensions to the basic neocybernetic model, it seems that there exist many alternatives. At least in simple environments, various structures can implement the necessary functionalities; and there exists a huge body of control engineering understanding readily available to boost the intuitions (for discrete-time control of dynamic systems, see, for example, [4]). Here, the above ideas are extended to tackle with the observed challenges.

Again, the idea following the cybernetic principles is to bring the world state back to intended balance state, or to the origin of the subjective variable system. An enhanced control structure is presented in Fig. 7.8. Still, this scheme is not quite universal: For example, here it is necessary that all goal points are balance points, so that zero error means zero control. There are three parts in this structure — modeling of change, prediction, and control construction — and some key points are briefly explained below.

First, it is *change* in the environment that is being modeled — what remains always constant is not interesting from the point of view of information acquisition or from the point of view of control: Variables that do not affect or that cannot be affected should be ignored. What is more, this change is defined as the difference between the actual state of the environment and the state that was predicted; this makes it possible to concentrate on phenomena that are truly new and contain the most of fresh information. When modeling the difference between the observed and the estimated state, one needs an additional signal coming “from the past”: The input/output structure becomes “two-directional”, input coming essentially from two sources (compare to Fig. 3.7). The past information is unalterable, thus not introducing additional dynamics in the model structure.

The middle part in the figure represents the key functionality, or the prediction of the future state based on the current state. The prediction is simple least-squares mapping between the former and the latter mental images — adaptation of this mapping model can only be carried out afterwards, when the future is visible, but the model can be used without such delay. To implement the least-squares mapping with minimum number of auxiliary structures, it is assumed that internal feedbacks in the neurons keep their activities at a certain level. When the coupling coefficients q_i are individually controlled to

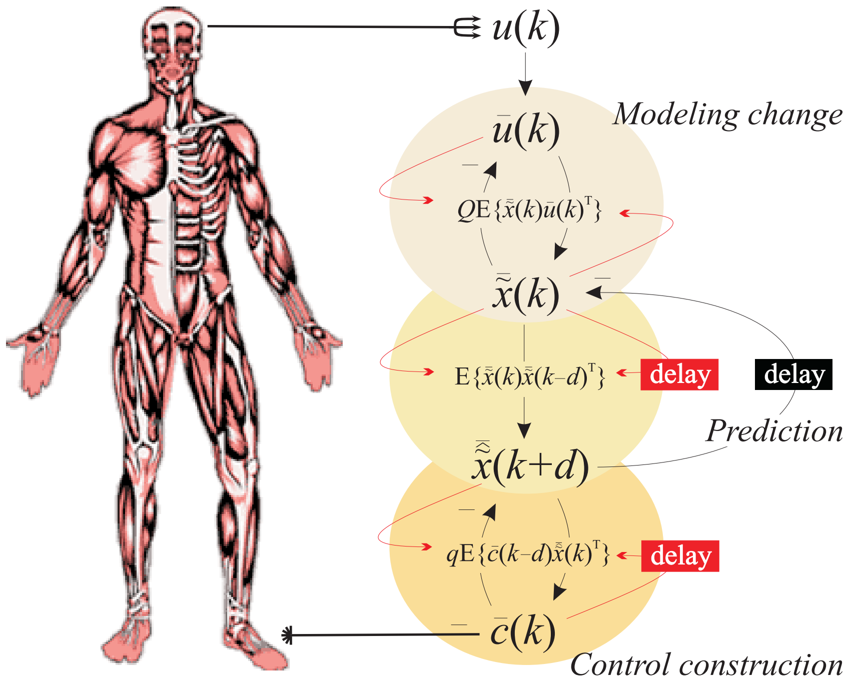

Figure 7.8: "Almost complete" solution to cybernetic control

make variance of $\bar{\bar{x}}(k)$ have value 1, then, according to the discussions in chapter 3, the whole covariance matrix $\mathrm{E}\{\bar{\bar{x}}(k)\bar{\bar{x}}(k)^T\}$ becomes an identity matrix, and the least-squares mapping from the previous to the next state becomes simply $\bar{\bar{x}}(k+d) \approx \mathrm{E}\{\bar{\bar{x}}(k+d)\bar{\bar{x}}(k)^T\}\bar{\bar{x}}(k)$. As all variations $\mathrm{E}\{\bar{\bar{x}}_i^2\}$ are equal, it turns out that triangularization of the covariance is necessary to distinguish between the variables. Again, modifications of the cybernetic adaptation strategy are needed: It is not the input and output that are used for updates, but this time it is the input and the *earlier input.*

Finally, the construction of control itself follows the same lines of thought as in the above simple control scheme in Sec. 7.2.2. Before further adaptation, appropriate initialization of the data structures is first needed: This means, for example, explicit stabilization of unstable systems. The control also needs to be bootstrapped: An external controller is needed in the beginning to instantiate the control model, and during adaptation, the cybernetic controller only gradually takes over.

The presented control scheme is versatile, at least in principle: If nonlinearity (the "cut" function) is included in the perception structures, one has *sparse-coded feature-based control.* The key point is that even the most complicated control tasks (like biped walking) can be implemented applying the piecewise linear structures that are based on local principal component models (see [34]).

The above discussions concerning cybernetic control are by no means exhaustive. The key point to observe here is that when trying to exploit the inherent time-domain structures of the world, extensions to the basic neocybernetic model structure are necessary. The above structures cannot be generalized to other domains — is there *something* universal that can be said about the world-induced cybernetic structures?

7.3 Planning and beyond

The above discussion concerning the cognitive system is by no means applicable to other cybernetic systems as such. In cognitive systems where the functionalities are based on neuronal connections one can easily design additional constructs that implement explicit prediction and other functionalities. In other domains there typically is less freedom, the functionalities being dictated by the physical laws of the environment. Whereas the cognitive system has evolved for planning, or for simulation of potential worlds and for consciously changing the environment to fit one's targets, in other domains there exist no such explicit goal-directedness. Or is there? It seems that it is difficult to reach some general model that would cover all cybernetic systems; some ideas are universal, though — and somehow addressing the intentional changes in the environment is one of such principles. Again, when looking at the very functional approaches that natural systems have found to tackle with such challenges, it is evident that there are lessons to be learned.

7.3.1 From reactivity to proactivity

When the cognitive system was taken as an example of cybernetic systems, some general aspects of the cybernetic models — like the possibilities and interpretations of sparse coded subspaces — could be made better comprehensible. But, after all, perhaps that example best illustrated how different the systems in different phenospheres can be. Whereas *intelligence* can be defined as the capability of tackling with and managing in new, unknown environments, *life* can be characterized as the capability of tackling with and managing in familiar, known environments. Intelligence is manifested in creativity, but life is manifested in routine. For some systems the changes in the environment repeat, and the future is known for a long time ahead. Mastering this routine, acting reasonably when one knows what there is to expect — this is the key challenge here from now on.

The case of cognitive systems illustrated the need to tackle with not only the current environment but also with the future environment. This is a new and crucial point: This far feedback control has been emphasized – it is what the cybernetic agents implement when they adapt to their environment, either explicitly (as in "intelligent agents") or implicitly (as in "selfish agents"). Feedback is a robust way to tackle with unknown environments, as the balance is efficiently restored after the disturbances are detected. But such feedback control is always reactive: You will do nothing before things go wrong. Prediction of the future disturbances would make it possible to implement *proactive control,* where disturbances are compensated before they ruin the system. When the sources of disturbances are known, one can implement — applying engineering terminology — *feedforward control.*

Feedback is the only reasonable control scheme when there is noise in the environment — or actions of other unknown systems. But when the environment is already well in control so that its degrees of freedom can freely be manipulated by the system (or set of systems), and when all degrees of freedom are under control, the problem setting is very different. The systems can take an active role. To begin with, the environment changes, and after that, the system changes; but when the systems are "mature" and dominant in their environment, it can be so that as the systems change, the environment follows. When the environment is thus under control, there is no limit: The histories of cumulating manipulations can become longer and longer. The environment can be tailored at will by the systems — but where should it be driven to? This is not important, the key point is that all subsystems agree upon the direction so that the balance can be maintained. As Lewis Carroll puts it:

> "Cheshire-Puss, would you tell me please,
> which way I ought to go from here?"
> "That depends a good deal on where you want to get to," said the Cat.
> "I don't much care where," said Alice.
> "Then it doesn't matter which way you go," said the Cat.

When the population is a model of the complex nonlinear world, it is the individual submodels that have to map the whole Wonderland. Only through such ant-like search in that world the ranges are found, the possibilities becoming investigated, so that the emergent-level model of local landscapes can be compiled (even though this knowledge will ever remain distributed). In retrospect, different route selections can be evaluated against each other, and then it is the Darwinian mechanism that can efficiently optimize among possibilities. When the world has been mapped, it is reasonable to follow the right path. The cognitive models can be assumed to construct their model of the future by trial and error; in other phenospheres, however, one possibly cannot afford mistakes, getting lost in the forest. Especially if the route is long, there is no time to waste. The cognitive system has its limitations — remember that even learning the muscular sequence of the golf swing takes a lifetime. In practice, to find the desired place again, to take only the right turns, one necessarily needs *instructions.* To implement such route maps, nature has been very innovative.

All complex population systems seem to be based on different kinds of instructions, and there are different kinds of implementations in different phenospheres. In biological systems these instructions are coded in the genome, being innate, and the flow of information is one-directional, so that the phenotypes cannot be reduced back to the underlying genotype. On the other hand, memetic systems are based on the cognitive "tabula rasa", the instructions getting acquired from canonical scriptures, and the flow of information is partly two-directional as it is the clever minds that produce the scriptures (even though this explication of expertise is typically difficult). The achievements in the cultural arena would not be possible if they were based merely on the generic adaptation capability of the cognitive medium — not everything can be learned the hard way. The fast advances in cultural evolution are only possible because the production of culture is cumulative, and the evolutionary developments there — creation of new cultural achievements — can directly be based on the prior ones.

Note that it is still the same optimality criterion as before that is assumed to guide evolution — the match with environment, or ability to exploit environmental resources determines fitness. Among the structures, however, there are no visible “gradients”, there is no visible direction to go, and enhancement can be observed only in retrospect. Thus, evolution of structures is not absolute in the sense that goodness can be compared only as related to alternative structures that have been experimented with; there is only “partial ordering” of locations on the map.

As was learned from the case of cognitive systems, trying to reach out from the current time towards the future necessarily requires structural developments in the system; the more ambitious one is, the more sophisticated structures are needed. Also the steps along the longer paths towards the desired locations in the future are structural changes. The steps of structure change are combined with parameter tuning in between; between the structural changes the balance is restored around the new structure — this way, not all details need to be codified in the instructions. Such succession of qualitative changes gradually modifying the system outlook are characteristic especially to evolutionary processes. Before, it was observed that catastrophes are the key to structural changes, the whole old structure being reset; however, explicit instructions are a way to avoid catastrophes, new structures being build upon the existing ones. Individual systems are not to question the instructions — it is the interplay between the systems and the overall environment that is the most important thing, the subsystems just supporting the emergence of something “better”. That is why, in some cases the death of the system is also predestinated in the instructions; this kind of *apoptosis* can take place when the system has done its share in changing the environment.

How is it possible that there is such wisdom built in the very mundane systems? An example is needed here, and, as it turns out, the case of biological systems is very illuminating: Following the above lines of thinking, the levels of individuals, populations, and whole ecosystems become the same, being based on individuals following the same instructions — just interpreting the instructions in different ways.

7.3.2 Ontogeny of systems

The process of finding balance in a system, as discussed before in linear terms, can in more complex systems be highly nonlinear, becoming a full-grown organism consisting of structural changes.

Development of a complex system is a step-by-step process. The system has to be bootstrapped: The lower-level subsystems first need to be instantiated, all attractors activated within a functioning environment — indeed, they have to be brought to life — before the higher-level systems can survive in that environment. A complex system cannot be instantiated as a one-step process – or, at least, nature has not found the way to do it. This means that a new individual has to repeat the same steps as its ancestors to become living. Yes, all steps from the beginning of life, starting from the simplest chemicals and catalysts, in principle have to be repeated.

There cannot exist structures of pure information; they must reincarnate in some physical form. And any physical system is vulnerable to decay and wear

— they must be regenerated periodically. This two-way nature of all systems gives rise to deaths and births of individuals, or “system carriers”.

According to the assumption of Ernst Haeckel (1834–1919) the development of an individual embryo repeats the development of the whole species, or, as he expressed it, “ontogeny recapitulates phylogeny” (see Fig. 7.9). Even though his idea has been heavily criticized, how else could it be? After all, as we now know, it *is* mostly the same genes that are shared by very different species. The same basic genes are shared by all of the biological living systems, even though these genes may be interpreted in different ways, and they can become activated at different stages of development. More complex life forms (that have assumedly emerged later) have newer genes of their own, but they still consist of the same underlying simpler functionalities, the whole path from the beginning to the end being covered in the “building instructions”. The more there is common in the two genomes, the longer history the species assumedly share. Of course, Haeckel’s idea is a simplification — essentially the same truths can be expressed, for example, in the form of von Baers Law: “Features common to all members of a major taxon of animals develop earlier in ontogeny than do features that distinguish subdivisions of the group”.

The system has to be instantiated in a single individual to become alive; for biological systems, this means individual animals, and for memetic systems, this means individual human minds. There is always a physical rack that is needed, and the system size cannot grow beyond the capacities of that medium. One concrete constraint is the life span: Because of the inertias in the environments, it takes time before the balance is reached after each structure change. For a system to evolve further, there must be enough time for the system to be “downloaded” — and this is only the basis where the new developments are to be built on. When the species history gets longer and more sophisticated, the instructions need to become more efficient; and it is not only the history of the one species but it is the history life on earth. It seems that the “higher” animals having longer history to repeat, have managed to streamline the development processes — in addition to typically having longer life times and duration of gravity in general. How can this be explained?

Sometimes processes become streamlined as shortcut paths are found between the original routes, the development becoming more straightforward. However, more typically, it seems that, at least to some extent, acceleration of code reading is built in the biological medium itself, and no structural changes are needed to boost the processes: Along time, balance periods between structural changes seem to become shorter and shorter. Where does the acceleration of processes come from? Remember that the steps in development are based on new genes becoming expressed, and these genes are there available, just waiting to become activated. This activation takes place whenever the level of appropriate excitatory factors has reached the threshold level; to make structural developments follow each other at a faster pace, it is only the question of making the underlying quantitative processes more prominent — typically this happens as the quantitative cybernetic matching processes are polished. Whereas the genes themselves are evolutionarily old and they are mostly shared by different species, it is the genetic control structures that have evolved later, making it possible to easily alter the details of gene expression. From the succession of waterfalls and

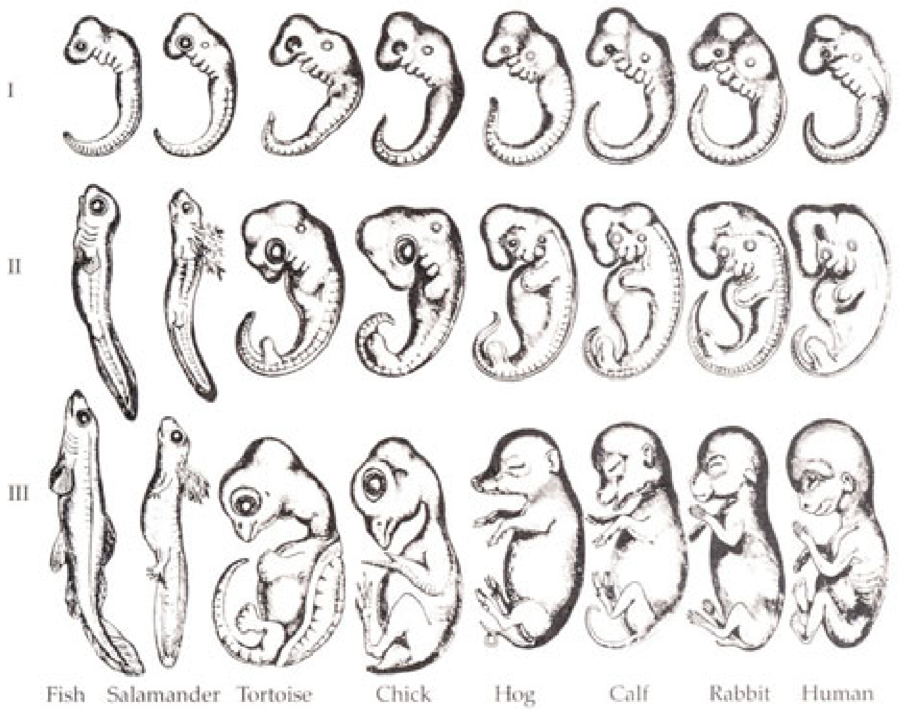

Figure 7.9: Graphs that are today usually called “Haeckel’s lie”. Ernst Haeckel claimed that the embryos of fish, chicken, human, etc., are evolving through the essentially same phases, repeating the common history of species development

quiet waters, structural changes and balance periods, the development processes seem to evolve towards torrents, continuous fast-flowing rapids.

The instructions need not be implemented strictly sequentially — as long as hierarchy among subsystems is maintained, the higher-level constructs being based on the lower-level ones. For example, in the developing embryo, the subprocesses are parallel and somewhat independent. The changes in expression rates of the corresponding genes can also develop at different rates. Indeed, such differences in gene expression properties are known in developmental biology as *heterocrony.* As the genetic controls become more efficient, control signals becoming stronger and more consistent, the genes are activated earlier; the faster some control starts the more prominent that structure typically is in the adult. Especially in vertebrates the basic structures are the same, differences in the outcome being to a large extent based on at what time during the development the genes started becoming expressed.

There are basically two main classes of systems: The biological ones in the chemical domain being based on genes are “natural systems”, whereas the memetic ones in the cognitive domain being based on memes can be called “man-made systems". There is very much in common among them what comes to the role of the "instructions". Just as genes are shared among species, memes are shared among cultural works, a new combination of memes being a structurally new "species", mental or physical artifact, to build further culture on. The "memetic phylogeny", or cultural evolution, is fast because it is free of the physical constraints of the interpretation machinery: The mental machinery is a universal medium for all kinds of memes, structural changes being implemented in "software". What comes to "memetic ontogeny", also the memetic systems need to be instantiated, starting from zero, in each mind separately. Again, the developmental subprocesses can be, if not parallel, still uncoordinated: When reconstructing a memetic system it does not matter in which order you read the books as long as you understand the used concepts. Similarly as the genetic systems, also memetic ones (like scientific theories) are streamlined as they are matured; to fit in a single mind, they need to be optimized to become extended. This streamlining does not apply only to the meme combinations themselves, but also to the medium: the mental machinery develops from the simple ways of thinking towards more mature ones. Indeed, the reasoning processes can also been seen as an evolutionary ones, starting from simpler mental structures and ending in more appropriate ones. In the beginning finding connections between mental variables is more or less random, new structural changes being called ideas or innovations, bursting out from prior balances to new conclusions of released tensions; but after rehearsal, such inference processes become more fluent and automated.

When modeling the most interesting cybernetic systems, it seems that mastering the evolutionary processes would be of paramount importance. To understand such phenomena one needs tailored frameworks to model sequential processes. It turns out that one needs domain-specific languages and grammars.

7.3.3 Representations of evolution

To understand living systems, evolutionary phenomena are perhaps the biggest challenge. All systems where there is evolution are basically based on sequential representations by nature. The reason for this seems to be that a sequential succession of instructions is nature's way of passing information over gaps between systems. Linear codes can be easily read and reproduced — copied, stored, and transmitted. Even though being sequential, implementation of such code cannot usually be characterized as being process-like: The interpretation of the code is detached from the time variable, instructions being read and substructures being defined in a somewhat sporadic manner. There do not exist strong mathematical tools to master such mappings between topologically so different structures. Still, conceptual tools for formalizing evolutionary non-continuous processes are needed: One needs compact model structures to capture the functioning of the codes. It is different kinds of *formalisms* and *languages* with special grammars and vocabularies that can be applied for capturing such codes. What do we know about such representations?

In the memetic domain, there exist ample evidence and experiences about the properties of codes and their interpretation. Contribution of AI or cognitive science in this context is that the memetic representations are studied a lot there, the connections between natural languages and the corresponding mental constructs being a central topic there. Perhaps some questions that have not even been formulated yet concerning evolutionary systems have already been answered?

The objective of all natural language use is to instantiate more or less independent subsystems in minds. How well that code matches the existing environment, how relevant it is, dictates how "living" that subsystem becomes, being perhaps later exploited for a larger-scale memetic system. A code-form representation should correspond to a high-dimensional associative representation, so that the knowledge can thus be stored outside the living system, making it possible be put alive in another mind later. The coding is not unique, and there are different kinds of codes, as the dynamics of the attractors can be waken up in different ways. In its most compact and explicit form, the bare bones of expertise can be represented as declarative rules that are explicit partial projections of the high-dimensional representation onto low dimension. There is plenty of material on the challenges of doing the inverse, when going back from the declarative to the associative, or from "novice" representations to expert representations. These age-old AI problems become a more general problem plaguing all cybernetic systems: The essence of a complex system is difficult to represent in code, and it is difficult to implement that code as a system. But if it is the nature's way of representing the system outside the system itself, the way to survive over the succession of deaths and births, it should assumedly be pursued also by humans trying to do the same.

As there is intuitively such a close connection between the codes in memetic and the genetic systems, one is tempted to speculate. Expertise is difficult to explicate, but it still can be written in books, no matter how fragmented that representation necessarily becomes — is it really so that nature has not found any way to reach such bi-directionality in the genetic system? The dynamic balances can be constructed using the genetic code, but it seems evident that the code cannot be modified by the system state. However, when comparing to the use of language, it is the available memetic codes that are recombined; language structurs need not be recreated, they are just activated appropriately. Similarly, perhaps the genetic code can be seen as a set of rather constant building blocks, gene atoms, and the main emphasis is on the instructions telling how to combine them. Just as texts can be constructed on the fly by combining the available memes in more or less fresh ways, gene functions can be combined in an equally flexible fashion. The epigenetic cellular state, or the vector of transcription factor activities, reveals how the control genes are reprogrammed. This system state can then be inherited in a Lamarckian fashion without affecting the underlying genetic codes. Perhaps the increased flexibility explains why the control genes seem to be so influential in higher life forms?

The mappings between the code and the functioning system are not one-to-one, not unique in either direction. It is clear that in the high-dimensional dynamic system of continuous-valued variables there is more information than can ever be stacked into a finite code consisting of distinct variables. But also in the opposite direction, when putting the code alive, misinterpretations are possible because the systems are instantiated in different environments — and it is, after all, the personal environment that determines the relevant attractors.

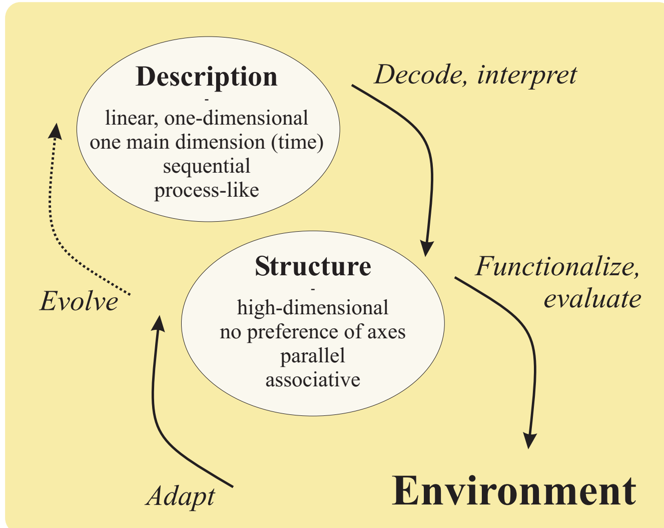

Figure 7.10: From sequential to parallel — each representation being meaningful and interpretable only in an appropriate environment

In Fig. 7.10, the relationships between the description of a system and its implementation is illustrated. The description determines the structure, and the system itself matches the structure against the environment by fitting its parameters — and simultaneously changing the environment. The system is the mirror of the environment only within the determined structural framework. It is the environment everywhere coupling things together, supplying for the interpretations: The environment is the key to assessing the relevance of a system as it is finally the match against the environment that determines the validity of a system and its internal tensions. Also the language of the codes, grammar and vocabulary, is determined by the environment, because it is the environment that has to interpret the code.

The role of the environment cannot be overemphasized, as it is the final judge supplying for the *domain-area semantics.* Only if the structures are interpreted in the right environment they can become living attractors; without interpretation all structures and signs are void. In this sense, one could speak of "natural semiotics". For example, the ancient texts are not only undecipherable sets of tokens, but they carry semantics even after the original culture is dead — assuming that the memetic attractors still exist in our culture, and the cultural context can be reconstructed.

The environment having such a dominant role, it is questionable whether there can exist any general theory of evolutionary systems. It seems that evolutionary processes cannot be abstracted away from the details of the properties of the underlying medium. A code is only meaningful if the environment — or the interpreter of the code — is also exactly defined. Interesting analysis is still possible — in the following chapter, ideas concerning such codes carrying the domain-oriented interpretations are studied in terms of an example case. There are so many efforts on trying to understand the memetic code, the natural language, so that, for a change, study the genetic code and the special challenges of the chemical domain. How does the “proteomic code”, the sequence of amino acids, dictate the protein structure?