Neocybernetics in Biological Systems

Heikki Hyötyniemi
heikki.hyotyniemi@tkk.fi

Helsinki University of Technology, Control Engineering Laboratory
Report 151
August 2006

Abstract

This report summarizes ten levels of abstraction that together span the continuum from the most elementary to the most general levels when modeling biological systems. It is shown how the *neocybernetic* principles can be seen as the key to reaching a holistic view of complex processes in general.

Preface

Concrete examples help to understand complex systems. In this report, the key point is to illustrate the basic mechanisms and properties of *neocybernetic* system models. Good visualizations are certainly needed.

It is biological systems, or living systems, that are perhaps the most characteristic examples of cybernetic systems. This intuition is extended here to *natural* systems in general — indeed, it is *all other* than man-made ones that seem to be cybernetic. The word "biological" in the title should be interpreted as "bio-logical" — referring to general studies of any living systems, independent of the *phenosphere.*

Starting from the concrete examples, connections to more abstract systems are found, and the discussions become more and more all-embracing in this text. However, the neocybernetic model framework still makes it possible to conceptually master the complexity.

There is more information about neocybernetics available in Internet — also this report is available there in electronic form:

`http://www.control.hut.fi/cybernetics`

National Technology Agency of Finland (TEKES) has provided funding during the research under several project frames, and this support is gratefully acknowledged.

This report is my "scientific legacy", and it is dedicated to my dear sons Elias and Matias, with whom our Adventures no longer can be more than mental ones.

On May 12, 2006 — anniversary of the Finnish Culture

Heikki Hyötyniemi

About the cover

Might is based on Wisdom. And the wisest man of all is Väinämöinen. But the Magic Boat is too large a system to master even for Väinämöinen. He needs to ask help of Antero Vipunen.

Antero Vipunen is the Earth Giant. Things grow on him and in his sleep he absorbs Natures secrets. — What happened then? For details, see Kalevala, the Finnish national epic.

Väinämöinen implementing "experiment design" with Antero Vipunen (graphics by Akseli Gallen-Kallela)

According to Kalevala, knowledge is power. Completely mastering a system is *being capable of presiding over its birth.*

In today's terminology, this all is about understanding the processes of emergence. The contribution of modern cybernetics is that *it may be the same formula to master all systems, big and small, living and man-made.* Knowledge is not only about understanding how systems work — it is about *making them exist* in the first place. — What does this mean? Please, read ahead.

Tunnenpa
systeemin synnyn.

Oleva tiedosta tehty
mielestä on ja mallista

Ajatus aineesta tehty
mitattavasta datasta

Contents

Level 0

Chaotic Complexity vs. *Homeostasis*

The deepest intuitions concerning real-life complex systems date back already to Heraclitus (about 540 B.C.):

- *Everything changes, everything remains the same.* Cells are replaced in an organ, staff changes in a company – still the *functions* and *essence* therein remain the same.
- *Everything is based on hidden tensions.* Species compete in ecology, companies in economy – the opposing tensions resulting in *balance* and *diversity.*
- *Everything is steered by all other things.* There is no centralized control in economy, or in the body – but the interactions result in *self-regulation* and *self-organization.*

However, after Heraclitus the mainstream philosophies developed in other directions: For example, Plato emphasized the eternal *ideas,* regarding the change as ugly and noninteresting. And still today, the modern approaches cannot satisfactorily answer (or even formulate) the Heraclitus' observations. What is the nature of the "stable attractors" characterizing complex systems? There have been no breakthroughs, and there will be no breakthroughs if the fundamental nature of complex systems is ignored. There now exists a wealth of novel conceptual tools available — perhaps it is the time to take another look.

0.1 Facing the new challenges

The ideas are intuitively appealing but they are vague, and there are many approaches to looking at them. Truly, it is a *constructivistic challenge* to try to explain a novel approach: Everybody already knows something about complex systems, and everybody has heard of cybernetics, but few people share the same views, and misunderstandings are unavoidable. That is why, there is need to briefly survey the history — or, as Gregory Bateson (1966) has put it:

> I think that cybernetics is the biggest bite out of the fruit of the Tree of Knowledge that mankind has taken in the last 2000 years. But most such bites out of the apple have proven to be rather indigestible – *usually for cybernetic reasons.*

0.1.1 Lure of cybernetics

The term *cybernetics* was coined by Norbert Wiener in 1948, when he published his book *Cybernetics, or Control and Communication in the Animal and the Machine* [86]. The underlying idea in cybernetics is the assumption that it is the dynamics caused by the interactions and feedback structures among actors that result in observed complex behaviors. According to one definition, cybernetics is the study of *systems and control in an abstracted sense.* Indeed, intuitively, cybernetics directly addresses the Heraclitus' challenge.

Because of its intuitiveness, cybernetics was thought to be a panacea — indeed, it was one of the first "isms" to become *hype.* Since its introduction, there has been a long history of false interpretations, not only in Western countries, but also in the East, where cybernetics was seen as (another!) "scientific" motivation for communism: *How to steer the society in an optimal way?* There still exists a wide spectrum of more or less appropriate connotations; perhaps the term is today mostly associated with "cyberspaces", and "cyborgs", or cybernetic organisms combining biological and non-biological organs. In biological and ecological sciences cybernetics gained a bad reputation as the hypotheses were too wild: Evolutionary processes simply do not take place on the level of systems.

Perhaps cybernetics is today free of incorrect associations. Anyhow, it turns out to be an excellent framework for combining control theory, information theory, and communication theory with application domains (biology, ecology, economy, etc.).

Cybernetics has already had its impact on today's scientific world. For example, being a framework for studies on clever interactions among agents cybernetics was one of the starting points beyond Artificial Intelligence. Similarly, as a framework of complicated feedback structures, cybernetics boosted the developments in the field of modern control theory.

However, the total potential of cybernetic ideas has not yet been fully exploited in control theory. At Wiener's time, control theory was still very classical, and even the most straightforward ideas sufficed, but during the years control field has considerably matured. It seems that the field of traditional, centralized control theory has by now been exhausted – it is time to implement the deepest cybernetic insights and *get distributed.*

0.1.2 Theories of complexity

In fact, cybernetics is just one view to understanding complex systems in general. There are also other approaches to attacking the challenges, all of them being basically based on the conviction: *Clearly, there exist similarities among complex systems.* Different kinds of intuitions are exploited, for example, in

[14], [38], [43], and [91].

Whereas today's control theory concentrates on the individual feedback loops, being too reductionistic, *general system theory* explicitly emphasizes *holism.* The original contribution in this field was given by Ludwig von Bertalanffy [9]. However, trying to attack all systems at the same time, the approaches easily become too holistic without concrete grounding. There is need to find ways to combine the wider perspectives with the concrete substance.

An opposite approach is to start from the bottom, from simple formulas or data, and hope for some order to emerge when some kind of manipulations are applied. This kind of "computationalism" is the mainstream approach in complexity theory today — indeed, the trust on the power of the increasing capacity of computers seems overwhelming: "In 20 years, computer will be more intelligent than a human"! Introduction to *computational biology* is given, for example, in [84]. But what if the iterations are chaotic, the results are sensitive to the initial conditions, and the simulations have no more correspondence with reality? Mindless thrashing of data only gives trash out. And the more challenging goal — how can computation make non-trivial phenomena emerge? How to sieve the *essence* out from the data?

The field of complex systems research is far from mature. No paradigmatic guidelines yet exist: There are no generally approved approaches, common concepts, methodologies and tools, typical application domains or problem settings. It has also been claimed that this "chaoplexity" is a form of *ironic science* where there are unsubstantiated promises, buzzwords, etc., more than there are hard results [40].

There are also more striking views. It has been claimed that the vagueness in the field is not due to the inadequacy of the theories, but we are facing the end of traditional science. For example, Stephen Wolfram who proposes the use of *cellular automata* for representing natural systems, proves that such a model family is too strong, and cannot be analyzed by traditional means [91]. From this he deduces that a New Science is needed — but he gives no hints of what that science would look like.

But there are also other ways to escape the deadlock: Perhaps the cellular automata was not the correct model family for representing complex systems after all. The unanalyzability is a property of the model, not of the system itself. If a more appropriate model structure is selected, perhaps old mathematics still works?

0.1.3 Return to basic mathematics

As observed by Eugene Wigner [87], in the past mathematics has been astonishingly efficient when explaining nature. Why should we be unlucky, why should it all end now?

It is clear that today's conceptual tools are insufficient when explaining the complex diversity. New concepts and structures need to be defined, and one needs an appropriate language for presenting and defining these concepts: As Ludwig Wittgenstein observed in his Tractatus: "What you cannot express, that you cannot think of". Wittgenstein spoke of natural languages — but it is mathe-

Syntax		**Semantics**
Scalar t	Free variable	Time, axis of evolution
Scalar i	Index	Agent identifier
Scalar $J(x,u)$	Positive-valued function	Cost criterion
Vector ϕ_i	Latent basis vector	"Forage profile"
Scalars q_i, γ	Adjustable parameters	(Inverse) "system impedance"; time axis contraction factor
Vectors x, u	States, latent variables; input signals	Agent (population) activities; set of resources
Matrices A, B	System matrices	Feedback interaction factors; interactions with environment
$\frac{dx}{dt} = -\gamma Ax + \gamma Bu$	Linear dynamic model	Matching with environment
$\bar{x}(u) = A^{-1}Bu = \phi^T u$	Asymptotic behavior	Dynamic equilibrium
$\mathrm{E}\left\{uu^T\right\} \approx \frac{1}{t}\int_{t_0}^{t} uu^T\, d\tau$	Covariance matrix	Mutual information structure
$\mathrm{E}\left\{uu^T\right\}\, \theta_i = \lambda_i \theta_i$	Eigenvectors θ_i; eigenvalues λ_i	Directions of information; corresponding relevances

Figure 1: Key symbols and definitions to be studied later. Simple mathematics, yes, but appropriate interpretations make a difference

matics that is the natural language of nature! Development of mathematics has always been directed by applications, so that the logical structures and concepts have evolved to appropriately and compactly describe real-life phenomena; and, when looking at complex systems, there are some special benefits:

- In mathematics, syntax and semantics are separated; it is possible to generalize and find analogues among systems.

- In mathematics, real numbers naturally capture fuzziness, non-crispness and continuity.

- In mathematics, parallelity of phenomena is transformed into high dimensionality, and there are efficient tools available for operating on high-dimensional data structures.

- In mathematics, time-bound phenomena, dynamics and inertia can efficiently be mastered and manipulated, and asymptotic behaviors can be captured.

And, of course, the clarity and unambiguity of mathematical expressions is invaluable — as compared to natural languages, this helps to avoid hand-waving.

It turns out that no new mathematics is needed to model complex systems, it is just new interpretations that are needed (see Fig. 1). Mastering some basic mathematical grammar is necessary: Specially, linear algebra and matrix calculus, and understanding of dynamic systems is essential. No New Science is needed, the Old Science still suffices — but as new interpretations are applied, there will be a New World, new ways of seeing the environment!

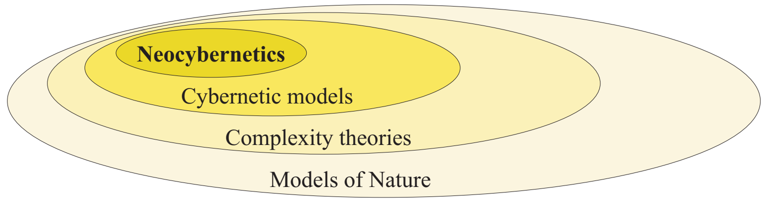


Figure 2: Neocybernetics offers a fresh view to studying real-life systems

0.2 Principles of neocybernetics

Cybernetics is a special view to look at complex systems, emphasizing dynamics induced by internal interactions and feedbacks. Further, *neocybernetics* is a special view to look at cybernetic systems (see Fig. 2). There always exist many ways where to proceed; it seems that neocybernetics combines mathematical compactness and expressional power in a consistent framework.

0.2.1 Capturing "emergence"

The key concept in complexity theory is *emergence* — some qualitatively new, unanticipated functionality pops up from accumulation of simple operations. There is a challenge here: If analysis of some higher-level phenomena cannot be reduced to analysis of their components, the traditional reductionistic modeling approaches collapse: The "whole" is more than the sum of the parts. This means that emergence is a somewhat notorious concept, emergent phenomena (like "life", "intelligence", or "consciousness") remaining outside the range of engineering-like "good" sciences. One could say that *emergent phenomenon is something that by definition defies definitions* — and what can you do then?

However, as emergence is indeed the *essence* of complex systems, it is necessary to attack this challenge. To reach good compact models, at each level one should employ the most appropriate concepts valid at that level — this means that emergence has to be "domesticated" somehow. The first objective here also is to *make emergence a well-defined, scientifically reasonable concept.*

When trying to formalize the idea of emergence, one can apply the very traditional modeling ideas: First, study explicit examples and construct an intuitive understanding of what the phenomenon is all about, and after that, find the common features and represent them in an explicit mathematical framework. And, indeed, there are many examples available where emergence is demonstrated in a very clear form. See Fig. 3, where the appropriate levels of abstraction are shown when modeling (gaseous) systems in different scales. Between each level, "emergence" takes place: Appropriate concepts, variables, and model structures change altogether.

> At the lowest level, it is the elementary particles that determine the properties of matter, the models on orbitals, etc., being stochastic. At the atom level, however, the Newtonian approximate ideal gas model with atoms as "billiard balls" becomes quite accurate, the appropriate concepts like velocities and moments being determinis-

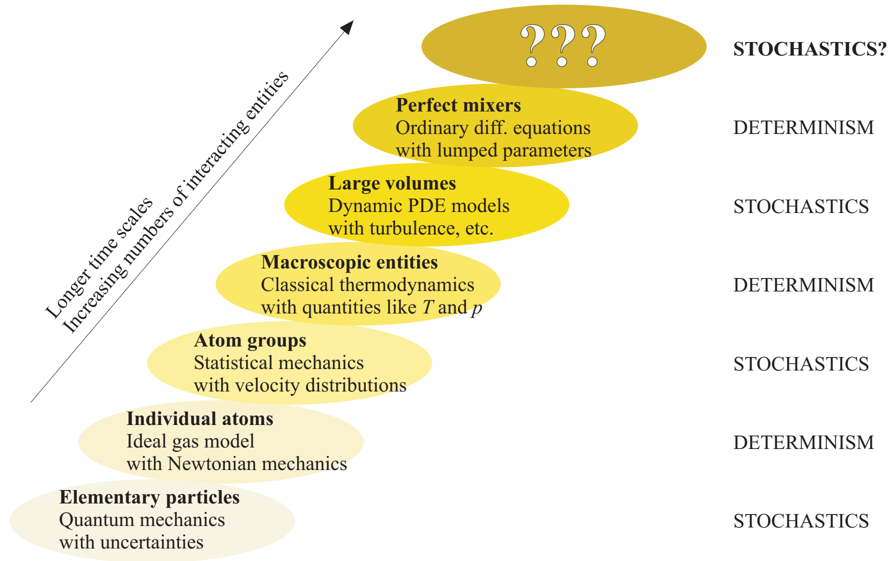


Figure 3: Different levels of abstraction are needed for modeling interactions of particles in different cases

tic. When there are millions of atoms, individual collisions cannot be tracked, and statistical mechanics becomes the modeling framework of choice. In still larger volumes, it is the deterministic macroscopic quantities like temperatures, pressures, and entropies that best characterize the system state. However, in still larger volumes, the temperature distributions cause convection and turbulence that can best be characterized in statistical terms. Assuming complete turbulence, the deterministic level of lumped parameters is again reached, where it is concentrations that only need to be studied.

Today, the level of deterministic first-principles models is already fully exploited. But to understand large systems consisting of such ideal mixers (like cells) one should reach for the still higher level of abstraction. Are there any lessons to be learned from the above hierarchy?

- First, it seems that one has stochastic and deterministic levels alternating in the hierarchy. Actually, this is no coincidence: For example, two successive deterministic levels could be "collapsed" into one.
- Second, it seems that on the higher levels the volumes (or number of constituents) is larger, and time scales become longer and longer.

How about exploiting the intuition on time scales: A higher level is reached, when the lower-level time scale is (locally) collapsed into a *singularity*, or when the time axis is abstracted away altogether. Note that at the higher (global) level, there can still exist time-related phenomena, so that still higher levels can further be defined.

Time axis is to be eliminated and individual signal realizations are to be ignored. Only the statistical properties are left there to characterize the overall signal properties; this must be done in an appropriate way, so that the properties relevant on the higher level are not compromised. It turns out that statistical cumulants like (co)variances, or expectation values of signal squares are beneficial in this respect.

Many problems fade away when the actual dynamic processes are abstracted using statistical (and static) system cumulants. But is this kind of ignoring of the time axis justified — when can such abstractions be carried out?

> To have statistical measures emerge, the signals have to be *stationary.* To have stationary signals, the underlying system essentially has to be *stable.*

However, out of all possible system models, the stable ones are rather rare: There must not exist a single unstable mode among the assumedly high number of dynamic modes. How could one assume stability in natural processes? — The motivation is simple: If the system were unstable, it would have ended in explosion (resulting in exhaustion of resources) or extinction already for a long time ago[1]. In this sense, one is not trying to model all mathematically possible systems — only the physically meaningful ones!

We are now ready to present the basic ideas beyond neocybernetic modeling.

0.2.2 Key ideas

The following principles can be used more or less as guidelines for deriving neocybernetic models, as will be demonstrated in subsequent chapters: Complex-looking phenomena are interpreted through the "neocybernetic eye-glasses". It needs to be emphasized that these principles are by no means self-evident (as becomes clear in Section 0.2.3): The proposed approach that has turned out to be advantageous is a result of iterative refinement processes, and, as it is always the case, the "highway through the jungle" without extra steps aside can be seen clearly only in retrospect.

Dynamic balance

In neocybernetic models, as presented above, the emphasis is on the final balance rather than on processes that finally lead there. In steady state one can directly attack the emergent pattern and forget about the details of complex nonlinear processes. It needs to be kept in mind that the balance here is *dynamic equilibrium,* or a balance between tensions, where external disturbances are compensated by some internal mechanisms. In practice, the compensating tensions are caused by negative feedback loops — but the implementation of these feedbacks is not of special interest, *as long as they can maintain the stability.*

[1] Stability here means *marginal stability,* that is, oscillatory systems, etc., are allowed — indeed, such marginally stable behaviors are typical to fully developed cybernetic systems. Truly, the key point is *stationarity:* The signals need to have statistically well-defined properties

Cellular systems have long been characterized in terms of balance or *homeostasis.* However, it needs to be noted that the concept of balance here is to be interpreted in a wider sense: The balance is defined with respect to only the selected variables. For example, it can be *derivatives* of some other quantities that are in balance, so that there is a balanced level of *dissipation* taking place in the system. Further, the system is assumed to be stable not necessarily locally but in a wider scale; for example, there can be oscillations as long as the system can maintain its integrity and the behaviors can be characterized in statistical terms. Dynamic transients are seen as secondary phenomena, being caused by natural strivings back towards balance after a disturbance.

Environment-orientedness

Neocybernetic systems are assumed to be explicitly oriented towards their environment, constituting "embedded systems" with their environments. The underlying intuition is that there cannot exist a cybernetic system in isolation. In this sense, the traditional system theoretic thinking collapses: a subsystem cannot be studied alone, without its connections to other systems.

This emphasis on the environment means, for example, that adaptation in the system is by no means a random process. A system reflects its environment, so that somehow it has to capture the properties of this environment. The available measurement information needs to be observed and stored in a reasonable way. All this means that it is not only the more or less random competition, "selection of the fittest", that is taking place in evolution — there are other, more consistent processes taking place, too, making such information gathering and storage more efficient.

Environment-orientedness gives another motivation for emphasis on balance: There is always scarcity of information, and the already existing structure has to be maintained while further information is gradually being acquired. Neocybernetic balance is a "kiln of emergent order". Only in stable conditions, when fast turbulent phenomena have ceased, something fragile can emerge.

High dimensionality

In practice, environment-orientedness changes to *data-orientedness.* No structure of the environment can be assumed to be known, only "measurements" of the environmental responses are available.

To make relevant information available to the modeling machinery, appropriate coding of information, or definition of *features* is needed. In neocybernetic models, *structural complexity* is substituted with *dimensional complexity,* that is, all possibly relevant features are simultaneously captured in the information structures (data vectors), hoping that the modeling machinery can construct appropriate connections among these pieces of data. Typically, the data are highly redundant, and new kinds of problems emerge. This means that efficient multivariate methods and corresponding mathematical tools are needed to analyze the neocybernetic models.

The features should capture the essence of the system; to reach this, the *domain-*

area semantics should somehow be coded in the variables — and, specially, as it was assumed above that the neocybernetic models are extremely environment-oriented, one is speaking of appropriate coding of *contextual semantics.* To reach this kind of coding, careful analysis needs to be carried out to capture the essence of the domain field in data structures. In [92], this bottom-up analysis was carried out for Hebbian neurons, whereas here it will be carried out for metabolic systems.

To reach the intended *universality* over the spectrum of all cybernetic systems despite the very different underlying realms, however, additional assumptions have to be made. If it is assumed that the mathematical model family is very constrained, so that indeed there are more systems to be modeled than there are available model structures, the behaviors of the different systems — within that model framework — must be analogous. How to determine such a model family that would be simple enough, still capturing the essence of systems?

Simplicity pursuit

The search for simplest possible representations is the traditional goal of practical modeling, being intuitively motivated by *Ockham's razor:* Simplest explanation is the "most correct". In a mathematical context, simplicity can often be interpreted as *linearity.*

The traditional approach to reach simpler analysis and manipulation of complex systems is to apply linear models. As a first approximation, linearity seems to offer rather good match with reality, at least if the nonlinearities are smooth and locally linearizable. The main benefit here is that for linear models one has extremely strong mathematical analysis tools available, no matter what is the system dimension; what is more, for linear model families one knows that the dynamic analogues work well, letting behavioral intuitions be transferred from a domain field to another.

In neocybernetic models, linearity is also taken as the starting point. The linearity assumption can be motivated in different ways:

1. In control engineering, it is well known that feedbacks "smoothen" nonlinearities. Specially in neocybernetic models where balance is emphasized, the deviations around the the equilibrium can be assumed to be small, and the transients can be assumed to have decayed, justifying the linear approximation.

2. High dimensionality typically makes it possible to find more linear models, at least if features are selected appropriately. For example, if the features include powers of signals, linear model can represent the terms in the Taylor expansion, approximating the nonlinearities.

But it is not only the pragmatic reasons — there also exist more fundamental motivations for taking linearity as the starting point. The belief here is that *there really exists a theory of cybernetic systems to look for* — and assuming that there will ever exist a general theory of cybernetic systems, *it must be based on essentially linear constructs.* There are no other alternatives — why? The system of cybernetic actors can be studied from outside in the top-down way, and in the bottom-up way :

1. **Top-down view.** Assume that a truly useful theory once is found. It is the linear models that are the only ones for which *scalability* applies, so that simple "toy world" examples can be extended to real-life scales — for large-scale nonlinear structures there cannot exist a general theory.
2. **Bottom-up view.** Assume that the complex system is to be based on identical underlying "agents" that do not share high-level strategies. It is only essentially linear combinations of underlying functionalities that can be implemented by such an unorchestrated bunch of competing actors — each of the actors only thinks for itself.

As it turns out later, the neocybernetic models are *optimal* in some very specific sense. If the "bootstrapping" in the underlying structures is carried out so that they are linear, the optimally adjusted layers later in the hierarchy of subsystems will also be linear. Linearity assumption is like a parallel axiom: It can be ignored, or it can be employed. In either case, a consistent non-trivial theoretical structure can be found.

The above reasoning only applies to "simple complex systems", and linearity is more like a guiding principle. The modeling strategy to be followed here is: Avoid introducing nonlinearities if it is not absolutely necessary, remembering that there always exist many alternative modeling approaches. Later, if extensions are necessary, the assumptions can be relaxed (such extensions are studied in the latter part of the report). Similarly, the originally static balance models can be extended towards dynamic models — but this should be done only after the basic nature of cybernetic systems is being captured.

0.2.3 Contrary intuitions

When comparing to traditional views of studying and modeling complex systems, the above neocybernetic starting points — views of emergence, role of time axis, balance, environment-orientedness, high dimensionality, and linearity — are very different, indeed contradictory.

What comes to *environment-orientedness* and *high dimensionality,* it seems that traditionally in chaos and complexity research, holism is studied in a very reductionistic ways. Typically, it is synthetic, isolated formulas that are iterated without connection to the environment, and it is hard to see how these "laboratory experiments" could be integrated in natural systems. The chaos theoretical models are extremely simple, often consisting of a single variable and a single formula. And also when explaining real-life complexity, the mysteries are often wiped under the carpet, into the twilight of the unknown: The issue of *emergence* has been "solved" by regressing it back to elementary levels. For example, Roger Penrose [63] claims that "cognition can be explained in terms of quantum-level phenomena". Indeed – there are always the underlying atoms and cells, etc., that implement the observed functionalities, but, as was explained above, these concepts are not the most economical way to express the higher-level phenomena. Similarly, it is individual chemicals that carry out the functionalities

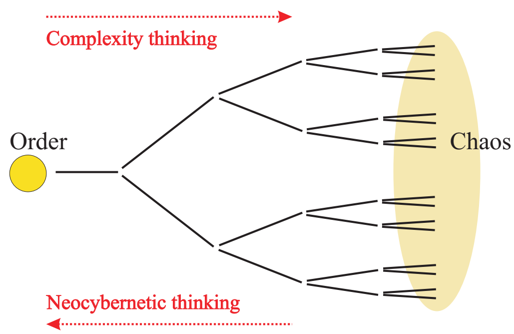


Figure 4: Complexity theory pessimism vs. neocybernetic optimism

of cells and organs; but, in the subsequent chapters, emergence is reached for by going *up,* not down.

The emphasis on *linearity* is perhaps the most radical assumption in neocybernetic studies, because nonlinearity is always taken as the starting point in studies of chaos and complexity theory — it is often thought that nonlinearity is the *essence* of a complex system. This nonlinearity view is well motivated, because theory says that linear systems are inferior to nonlinear ones: Without nonlinearity qualitatively new phenomena cannot emerge, and without nonlinearity there cannot exist chaos. But in linear systems there still can exist complexity, and, specially, there can be emergence of order. As it turns out, linear structures have not been fully exploited — or, rather, not all *interpretations* of linear models have yet been studied.

The intuitions concerning *balance* is a longer story. The idea of homeostasis has long history, indeed dating back to ancient times, but today it is regarded as a too poor starting point. Such views are formulated, for example, by Erwin Schrödinger [69] and Ilya Prigogine [64]: *The essence of life is in dissipative, non-equilibrium processes.* Static balance, steady state, is thought to mean death — interesting systems are seen to be extremely unstable, always being at the "edge of chaos". Whereas ordered state is uninteresting and complete disorder is uninteresting, it is the boundary line between order and chaos that is regarded as being of relevance. However, such boundary lines have zero length (in mathematical terms), their probability is practically zero, especially as such boundary phenomena are assumed to be unstable with their exploding Liapunov exponents. In neocybernetic models, however, things have to be studied from a different point of view: First, it has to be remembered that static and dynamic balances are very different things, the virtual placidity of dynamic equilibria hiding the underlying turmoil. Second, as it turns out, the balances are *stable* — this means that cybernetic systems are not at all as rare or fragile as the chaos theoretical instability-oriented thinking would suggest.

Abstracting the time axis away contradicts traditional intuitions about dynamic and turbulent nature of complex systems. It is the causally structured, or even *algorithmic* view to phenomena that rules today: Complex systems are seen as being composed of sequential processes. Individual one-at-a-time (inter)actions and explicit time structures are emphasized — no doubt because such action/reaction structures are easy to grasp. Another reason for the domination of such *process view* is, of course, the role of computer programs as the main tool in the simulation analyses. For example, it is *agents* that are seen as

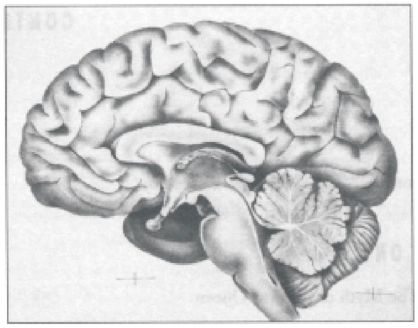


Diagram of the human brain *(Courtesy of Mittermeier)*

Map of Hamburg, circa 1850 *(Courtesy of Princeton Architectural Press)*

Figure 5: Do surface patterns reveal underlying similarities? (Adopted from [43]

the basic constructs in many complex environments, like in intelligent systems, and these agents are software constructs. On the other hand, if following the chaos theory paradigm, iteration is regarded as the paradigmatic route to complexity. The "butterfly effect" has been seen as being characteristic to complex systems, meaning that their behaviors cannot be predicted — indeed, models for them are more or less useless. In the neocybernetic setting, on the other hand, new hope is perhaps given to those who are struggling with modeling of complex systems. Because it is stable attractors that characterize the structures in cybernetic systems, there can exist consistent convergence even from differing initial conditions. Modeling *is* possible after all (see Fig. 4).

But one always has to be aware of the dynamic nature of the real systems, and the static models are dynamic ones in equilibrium. What is more, as it turns out, neocybernetic modeling itself is about balancing between dynamic and static worlds: Every now and then the static structure needs to be relaxed to escape the constraints of traditional thinking.

As explained by Herbert Simon [72], phenomena can be represented in terms of such processes or in terms of *patterns.* In neocybernetics, it is this pattern view that is pursued; these patterns are determined using statistical measures. And, further — it is not *surface patterns* but it is *deep structures,* or underlying latent patterns where the system is aiming at, when following its natural aspirations. The difference between surface patterns and deep structures needs to be emphasized: Also the traditional complexity theory is driven by patterns — Mandelbrot's fractals, Wolfram's sea shells, Kohonen's maps — but these visible formations, even though being intuitively appealing, do *not* capture the essence of systems (see Fig. 5). As Alan Turing has put it: “The zebra stripes are simple — I am more concerned of the horse behind”.

Summarizing, it can almost be said that the neocybernetic model is a *model of inverse thinking.* As it turns out, the relationships are “pancausal” rather than unidirectional; it is freedoms rather than constraints that are modeled, etc. Some additional insights are given below.

0.2.4 Neocybernetics in a nut shell

The starting points of neocybernetic modeling — linearity and balance, etc. — do not sound very intriguing. However, it is the strong mathematical tools, when letting their effects cumulate, that provide for nontrivial model properties. It needs to be noted that the presented approaches make it possible to define various consistent model families — the presented one, however, is claimed to the simplest one still giving nontrivial results. The following characterizations are studied in more detail when reaching higher levels.

The complexity theory being full of unsubstantiated promises, it turns out that *neocybernetics is the theory that puts the pieces (at least some of them) together.*

The neocybernetic model is a framework for studying variations, changes and tensions instead of immediately visible static structures. Counterintuitively, this analysis of variations is reached through the analysis of balances.

The role of dynamic balances is crucial when constructing neocybernetic models — indeed, the emergent patterns that are modeled are “structures of stability”. The neocybernetic model is a *model of balances,* or, if put in a more accurate way, it is a *balanced model of balances* (or *higher-order balance*) taking into account the properties of the environment, as determined by the statistical signal properties. The neocybernetic model is a map of the relevant behaviors corresponding to the observed environment, determining the behavioral spectrum of the system, where “behavior” means reactions to environmental excitations.

In a nonlinear system, uniqueness of the balance cannot be assumed; indeed, the neocybernetic model covers the spectrum of alternatives or potential balances, as determined by the environment. The neocybernetic model is a *model over the local minima* rather than a model of the global optimum, assuming that an appropriate cost criterion is defined. Traditionally, the single global optimum is searched for in analysis and in design: This results in theoretical deadlocks (compare to NP problems [73] — finding a large number of suboptimal solutions is typically much simpler than finding the absolute optimum). Also nature has no centralized master mind; it is facing the same optimization problems, seldom finding the strictly optimal solution: In this sense, the model over the local minima better captures the possible alternatives and essence (remember Heraclitus: “You cannot step in the same river twice”).

Because the system is optimized in a certain sense, the representations are (more or less) unique. The neocybernetic model is a “mirror image” of its environment, being itself a model of the environment, capturing relevant behavioral patterns as manifested in data. There exists certain kinds of *symmetries* between the original image and its model. This property makes it possible to draw conclusions, for example, about such high-spirited concepts as *intersubjectivity* and *interobjectivity.*

Because of the simple structure of the models, intuitions can efficiently be exploited: For example, the idea of analogues can be extended to partial differential equation models. The neocybernetic model can be seen as an *elastic system,* where the internal tensions compensate the external forces. The deformations are proportional to the forces (behaving like a steel plate) whatever is their physical manifestation. The electrical analogue makes it possible to conceptually manage neighboring cybernetic systems: There is maximum power transfer among the systems when they are matched so that their input and output impedances are equal.

There are close connections to today's research activities: Neocybernetics gives a framework for *distributed agents and networks* where there is no centralized control. It may also offer a framework for data-based modeling approaches and *computationalism.*

The negative feedbacks constructed in the neocybernetic model are control structures. The different dynamic equilibria result from changing inputs, or "reference signals" – thus the neocybernetic model is a *model-based adaptive controller* trying to compensate the disturbances coming from the environment. Further, this can be extended: The neocybernetic model is a *means of reaching maximum entropy* (or "heat death") of the environment. This means that the modeling framework offers a means of attacking the problems of cumulating improbability, and even for *inverting the arrow of entropy.* For example, it can be said that *life is a higher-order dynamic balance in some phenosphere.*

Indeed, neocybernetics offers tools for understanding the *whirls in the flow of dissipation.* These stable attractors are information-determined structures crystallizing the dependency structures observed in the environment. These intuitions can be applied to many very different domains from biological systems to cognitive ones, even to the Theory of Mind.

As it turns out, many of the neocybernetic issues have a more or less philosophical dimension. Without concrete grounding, such discussions are hollow and void, and they lack credibility. It is necessary first to define the concepts — or "whirls in the infosphere" — and this will be done next, the application domain being that of living cells.

— How to read the subsequent texts? Different chapters characterize specific aspects of cybernetic systems from different points of view, and they are best suited for people with different backgrounds. Together they are intended to form "ladders" towards understanding the steps in evolution — for changing such discussions into real science, or, indeed, into natural philosophy (see Fig. 6). The chapters marked in blue are mathematically involved or contain detailed physics or chemistry. The red chapters are more philosophically oriented. On the left-hand side, there are the *analyses,* being based on observations, whereas on the right-hand side, there are the *syntheses,* starting from first principles. The main line of thought in the middle tries to draw balanced conclusions between the tensions. The material is not self-contained, though: In the beginning, one should (in principle) get acquainted with *complex systems theory, multivariate statistics, artificial intelligence, biochemistry, ...*

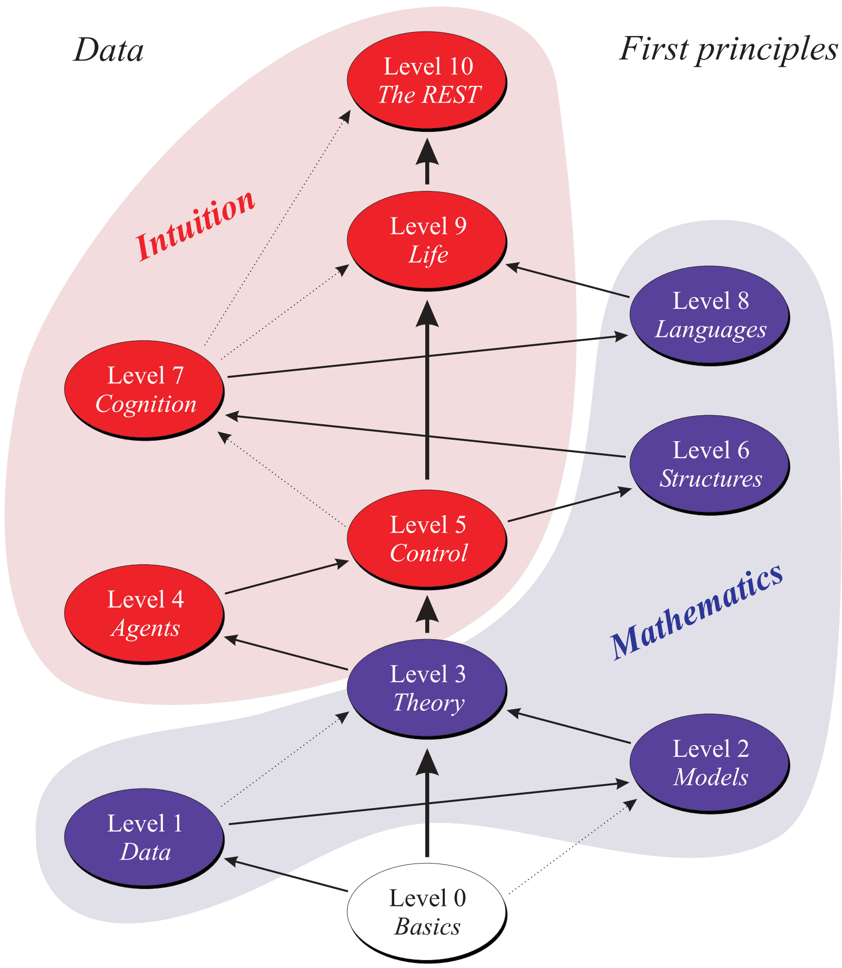


Figure 6: The chapters presented as ladders towards understanding of neocybernetic systems and emergence in them (sorry, but without some mathematical concepts there is *no route* to intuition)

Part I

Basic Models and Interpretations

Level 1

Genomics, Metabolomics, and *Distributed Networks*

There are very different kinds of subsystems in a living organism. To reach the level of *systemic biology,* one needs to be able to combine different model structures. The great promises have encouraged the researchers already for a long time to try and construct combined models (for example, see [85]).

When constructing models, it is first necessary to capture the essence of a domain field in mathematical structures. This is a delicate challenge, and domain-area expertise is needed. However, it seems that when one follows the neocybernetic guidelines, homogeneous representations can be defined for different kinds of systems. In this chapter, appropriate representations for information are first defined, starting from concrete low-level models, and abstracting them towards general model structures. Further analyses in subsequent chapters are based exclusively on these data representations.

1.1 Experiences with "artificial cells"

Information of the biological processes has increased immensely: Capability of reading the genetic code, and new ways of gaining information of the genetic activities (like *Chromatin Immunoprecipitation* or ChIP technique, see [74]) has delivered us large amounts of data. It has been assumed that being capable of deciphering the genetic code is enough to reach the higher level of understanding of what life is. Perhaps some day all dependencies between phenomena and control structures within a cell have been found. Is this not the ultimate goal?

Unfortunately, this is just one step towards capturing the essence of life processes, and the goal will *never be reached this way.*

What is the contribution an engineer having no background in biology can have, giving advice to domain area experts? — Counterintuitively, the engineer's contribution can offer wider views here. Systems thinking is universal, and, experiences from other fields can be exploited. Perhaps the same deadlocks can be avoided?

Understanding of complex systems is the challenge also in industrial automation systems. Despite the detailed system models, computers and simulators, the behaviors and qualitative properties of the overall system are becoming more and more difficult to understand. The systems are becoming like *artificial cells* themselves:

> Industrial plants also have *metabolism,* raw materials being exhausted and others being produced. Originally, the production can be far from optimum, but as soon as dependencies among variables are recognized, they can be used for constructing new feedback structures to implement more efficient and robust production. However, as the complexity of control structures cumulates, the system-level properties cannot any more be easily seen — even though all individual control structures are explicitly known, indeed, even though they have been explicitly designed and optimized.

In both cases, natural and man-made cells alike, it turns out that the goal of "evolution" is overall efficiency of production, no matter whether it is humans that are acting as agents for development or not. This can be reached by implementing mechanisms for reaching best possible production conditions; and this system integrity needs to be maintained without collapses. To maintain such balance, the system has to respond appropriately to the spectrum of disturbances coming from the environment.

Mastering huge amounts of data and finding "holistic understanding" out from it — this is the common goal in both cases, in artificial and natural cells alike. And, indeed, here it is the engineering tools and intuitions that can be exploited to find new kinds of approaches for attacking the complexity.

So, how to see the forest for the trees — how to see the cell metabolism for the chemical reactions, or, further, how to see the organ functions for the cellular phenomena? First, the details of the systems need to be understood and captured in data.

1.2 Modeling cellular processes

When aiming towards truly adapting systems, the model structures should not be fixed beforehand. It has to be assumed that there is minimum number of preprogramming, and the final structures have to be extracted directly from the observation data. One needs strong consistent frameworks where the observations can be interpreted. The hints of structure have to be coded in the data, and the mathematical machinery has to be capable of exploiting these hints. Indeed, this is a very ambitious goal.

1.2.1 From formulas to behaviors

Traditionally, modeling of complex systems is like search for the philosopher's stone: One tries to find the magical formula that explains all behaviors. Indeed — all systems, large and small, are assumed to be governed by underlying

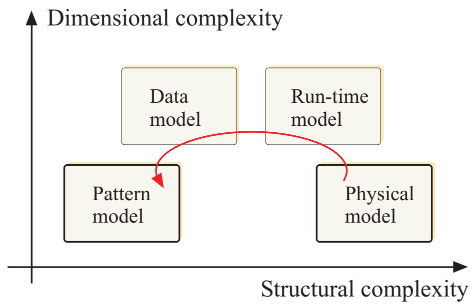


Figure 1.1: Reaching structural and dimensional simplicity at the same time

formulas, the complexity typically being manifested in nonlinearity; this kind of thinking is reflected also in today's approaches in complexity theory. However, in the case of truly large systems this objective becomes obsolete and absurd: As the systems are distributed, their behaviors cannot be compressed into some compact kernel.

Traditionally one tries to find the simplest possible formula that describes the system as isolated from its environment and other systems. The system alone is thought to represent itself in the most accurate way. But natural systems are never alone. In line with the neocybernetic assumption of environment-orientedness, it is assumed that it is the environment of the system that determines how the system is run and which of the potential behaviors become actual, selecting the subset of possible behaviors that is to be excited. In a sense, the environment carries out experimenting with the system, changing the conditions, and the system responds, finding a new balance reflecting the properties of the coupling between the system and its environment.

To unambiguously characterize the system behaviors, also the properties of its environment need to be quantified. In a complex environment, the simplicity goal cannot any more be reached as there exist a multitude of variables determining the environment. One is facing a problem of complexity exploding in two ends: Instead of simplicity, there is structural complexity in the form of the system model, and there is the dimensional complexity in the form of environmental variables.

However, as shown in Fig. 1.1, the structural complexity of the model can be ripped off by letting the complex behaviors be "interpreted" by the environment. When the environment is seen as a "simulator", and when one records the resulting behaviors, one has only homogeneous, bare numbers left. There is a high number of this kind of structureless data; assuming that the data is collected in an appropriate way, the relevant behaviors are present in that data, yet in a highly redundant form. The remaining task is that of detecting the patterns, finding the compressed representation of the information buried in the data — if this can be accomplished, one has a representation of the system that is simple in terms of structure and dimension at the same time.

How is such simplification possible, and how could it ever be done without human control and specific domain-area knowledge? First, the structural complexity in the form of nonlinearities can be changed into dimensional complexity when different kinds of nonlinearity prototypes are included in the data. Among the multitude of simple nonlinear features, the original nonlinearity can be approximated as a weighted combination of simpler ones. The question what these domain-specific nonlinearity prototypes are and how they can be isolated by applying appropriate data preprocessing so that model structure itself can be kept simple and general is discussed in this chapter, concentrating on the domains that are characteristic to a biological cell. After this, when the "behavioral essence" is present in the data, the second phase becomes possible: That of compressing the internal structure in the data and crystallizing the behavioral features. Only by making some assumptions about the nature of information (see chapter 2), the compression of the high number of environmental and system-specific data is possible.

The complexity becomes parameterized; the more there are features, the more accurately the behaviors can be captured. One can question whether the original function form can be represented by a set of other functions — but from the modeling point of view, if there is no difference in behaviors, the implementation has no relevance. According to the *identity of indiscernibles* originally due to Leibniz one can even claim that, "if all attributes of items A and B are identical, A and B are the same". Here data represents structure, the observations being assumed to capture the identity of the object, relevance of phenomena is defined in terms of visibility in data: As interactions make a difference, everything that is of relevance must be observable. This has only pragmatic motivation — but, later, closer look in these issues is taken: After all, the natural systems also only see the data available in their environments, and they try to tackle with it.

High dimensionality is easier to tackle with than different kinds of complicated structures — it turns out that the same tools are applicable to all systems. When the complexity has been transformed into high dimensionality, one can characterize the modeling task as being a *search problem*: When determining the model, or the patterns characterizing the data, one is facing a technical problem of finding the location in the parameter/variable space that corresponds to the observed behaviors. As the dimension grows, the search space grows exponentially. However, in the appropriate mathematical framework this search can be implemented efficiently in a *parallel* form. In (linear) vector spaces, it is multivariate statistical methods based on linear algebra that turn out to be efficient tools (see chapter 2).

The delicate relationship between the system and its environment is studied later closer; here it is enough to observe that all relevant nuances are represented in the data. Data selection essentially affects the modeling results, and selecting reasonable data and preprocessing it appropriately is a key question. One needs to find a coding where the system state can be captured in data; optimality of the representation needs not yet be worried about. To reach this, something needs to be known about the system structure — here it is assumed that what one is looking for is representations of *networks*.

1.2.2 Approaches to networks

It seems that an appropriate framework for studying the spectrum of the distributed cellular processes is that of networks, consisting of more or less independent interacting actors. The genetic system constitutes a network, where genes regulate each other, and also the metabolic system can be seen as a network among chemicals. And other complex systems, too — like ecosystems — are networks between individuals. How to model such networks? There exist dozens of alternative approaches — let us study some examples from opposite ends of the continuum.

Graph theory

The traditional approach to modeling networks, dating back to Leonhard Euler, was graph theoretic: The connections between nodes are assumed to be “crisp” — either there is an arc or there is not. Causal structures can be modeled using directed graphs with unidirectional arcs. This is also the traditional approach, for example, when representing control relations among genes, or when representing metabolic cycles.

However, such graphs are *descriptive* as models — easy to grasp but difficult to apply. There is the same problem as there is with all qualitative models: Just knowing that there is some connection is not enough. Altering the threshold level, the network easily changes from sparsely connected to more or less fully connected. And as is known in control engineering, the numeric values in the feedback structures essentially determine the properties of the whole system.

Indeed, in real networks, there is a continuum of interaction effects: The connections are not of “all-or-nothing” type. The graph models can be extended by adding weights to the arcs, etc., but at some point it is better to rethink the structure all over.

Probabilistic networks

Probability theory offers consistent ways for defining weighted arcs. In Bayesian networks, the theory of conditional probabilities is exploited, and chains of “evidence” and resulting probabilities are expressed as tree structures. Whereas graphs suffer from weak mathematical theory, Bayesian networks benefit from strong underlying mathematics – assuming that the assumptions hold: The evidence have to be independent of each other, etc. However, the nodes in real life are often not independent of each other. There exist feedback loops and alternative paths in complex networks, making the conditional structure intractable, or at least very complex (see [62]).

Another popular model family for capturing dynamic phenomena is based on Markov models, where the state transitions are probabilistic. However, now the problems are caused by the dynamic structure: It is difficult to find the typically large number of free parameters in such models. In real life, only a too narrow view of the potential dynamics is seen to identify the parameters; experiments typically are not very *persistently exciting.* And, again, there is the basic problem: The causality structure should be explicitly resolved to reach useful models.

Neocybernetic approach

After all, practically every node in a complex network is connected to all others, either directly or through intermediate steps. The neocybernetic approach is explicitly opposite to the traditional interaction-at-a-time analyses: Now, *pan-causality* is taken as the starting point. It is assumed that, after all, all nodes are simultaneously causes and all are effects, with the exception of the explicit system inputs. The network becomes more or less fully connected.

In balance, after the transients have decayed after some disturbance, the causal effects find their balance of tensions, assuming — in the neocybernetic spirit — that the underlying interactions and feedbacks are capable of maintaining the balance. It does not matter what the details of this stabilization are as long as the balance always is finally found.

The neocybernetic model tries to avoid the above problems of traditional network models. It is a numeric model, consisting of non-crisp connections, but the numeric values of these connections are not determined applying some probabilistic hypotheses, but by observing the relations between the materialized node activities. There is no centralized control or explicit network structure assumed or solved, so that there is no need to determine for the individual interaction strengths. Indeed, as is well known, the substructures in a closed loop system cannot be distinguished — this truth is implicitly accepted in the neocybernetic framework.

If the actual causality structures and dynamics are ignored, one could wonder, what is there left of the system? The neocybernetic model is a static balance model, or actually a model over the spectrum of balances as the environmental inputs change. If the system happens to be linear, the system state is unique, being a linear function of the input. These issues will be studied in more detail later, and they will be exploited accordingly. Next it will shown that, truly, the dependencies among variables in biological networks can be assumed (locally) linear.

1.3 Case 1: Metabolic systems

As presented in [92], the original starting point in neocybernetic studies was analysis of *Hebbian neurons* — however, when modeling biological systems, such analyses do not deliver useful information. It is *proteins,* for example, that are the means for implementing the cellular and organ-specific behaviors: A more relevant framework in this case is that of *organic chemistry.* It is the metabolic processes that eventually determine the cell behaviors, being the visible manifestations of the cell character.

Study a hypothetical example reaction, where there are α reactants on the left hand side, being denoted as A_i, $1 \leq i \leq \alpha$, and the β products on the right hand side are B_j, $1 \leq j \leq \beta$:

$$a_1\mathrm{A}_1 + \cdots + a_\alpha\mathrm{A}_\alpha \quad \underset{k_\mathrm{A}}{\overset{k_\mathrm{B}}{\Leftrightarrow}} \quad b_1\mathrm{B}_1 + \cdots + b_\beta\mathrm{B}_\beta, \quad \Delta H. \tag{1.1}$$

The metabolic processes are typically reversible, so that the reaction can take

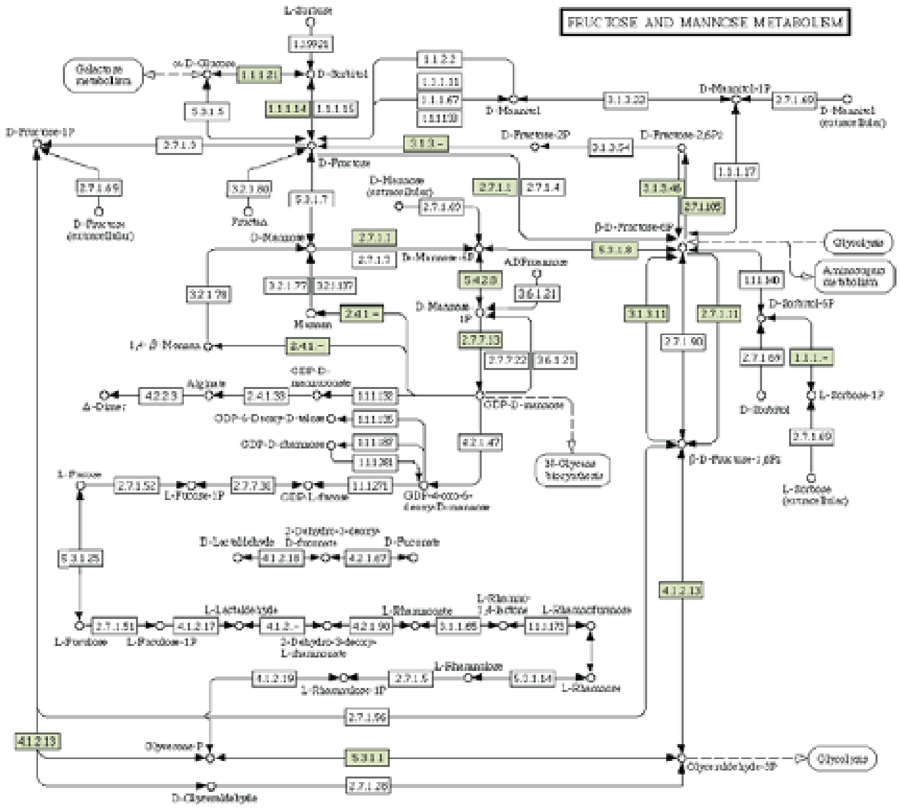


Figure 1.2: An example of metabolic pathways (courtesy of MediCel)

place in both directions (k_B being the reaction speed in forward and k_A in backward direction). Symbol ΔH denotes the change in enthalpy, or inner energy, when the reaction takes place. It needs to be recognized that it is not only chemical reactions that can be expressed using such formulas; also phase transitions, etc., can be expressed in this form.

However, chemical processes in metabolic systems can be very complex. For example, the active reaction chains in yeast when mannose production is taking place is shown in Fig. 1.2. And, what is more, such chains are just a part of the story: There exist overlapping sub-networks, and depending of the “projection”, the outlook of the graph changes. If some connection is explicitly cut, for example, applying some gene kick-off technique, the results are typically not what one would expect. The cytoplasm is typically strongly buffered — there seem to exist reserve mechanisms for compensating for the disturbances.

One needs a mathematically more compact representation for chemical reactions. How to “cybernetize” chemical reaction models applying the neocybernetic principles?

1.3.1 Applying the neocybernetic guidelines

In short, the goal here is to capture the domain area semantics, or expert knowledge in distinct pieces of information, and then pack this information into a compact form that makes it possible to apply the mathematical machinery. What is this domain-area semantics to be coded, then? From the point of view, the key point is to somehow capture the balances in the system.

Information representation

The first problem is how to represent such a chemical reaction formula in a useful numeric form. It seems that a practical way to code the reactions in a mathematically applicable form is to employ vector formulation: Define a vector C containing all chemical concentrations so that all A_i and B_j are represented there among the elements. The "chemical state" can assumedly be captured in this vector.

Let us look how this vector presentation can be exploited. If the coefficients $-a_i$ and b_j from (1.1) corresponding to the chemicals are collected in the vector *Gamma*, one can express the total concentration changes in the system as

$$\Delta C = \Gamma\, x. \tag{1.2}$$

Here, x is a scalar that reveals "how much" (and in which direction) that reaction has proceeded. When there are many simultaneous reactions taking place, there are various vectors Γ_i; the weighted sum of reaction vectors Γ_i reveals the total changes in chemical contents (assuming that the vectors are compatible).

Using the above framework, metabolic systems can in principle be modeled: If one knows the rates of reactions, or the scalars x_i, the changes in the chemical contents can be determined. This idea of *invariances* within a chemical system have been widely applied for metabolic modeling; the key term here is *flux balance analysis (FBA)* (for example, see [27]). However, the rates x are not known beforehand, and, what is more, the reactions are typically not exactly known.

In many ways, the model structure (1.2) is not yet what one is looking for. The main problem there is that the flux balances only capture the *stoichiometric,* more or less *formal balance* among chemicals. It does not capture the *dynamic balance,* whether or not the reactions actually take place or not. Luckily, there exist also other ways to represent the chemical realm.

Thermodynamic balance

There is a big difference between what is *possible* and what is *probable,* that is, even though something may happen in principle, it will not actually happen. To understand the dynamic balance, the reaction mechanisms need to be studied closer.

Assume that it takes a_1 molecules of A_1, a_2 molecules of A_2, etc., according to (1.1), for one unit reaction to take place. This means that all these molecules have to be located sufficiently near to each other at some time instant for the forward reaction to take place. The probability for one molecule to be within the required range is proportional to the number of such molecules in a volume unit; this molecular density is revealed by concentration (when the unit is mole/liter; by definition one mole always contains $6.022 \cdot 10^{23}$ particles). Assuming that the locations of the molecules are independent of each other, the probability for several of them being found within the range is proportional to the product of their concentrations. On the other hand, the reverse reaction probability is proportional to the concentrations of the right-hand-side molecules. Collected together, the rate of change for the concentration of the chemical A_1, for example, can be expressed as a difference between the backward reaction and forward reaction rates:

$$\frac{d\,C_{\mathrm{A}_1}}{d\,t} = -k_{\mathrm{B}} C_{\mathrm{A}_1}^{a_1} \cdots C_{\mathrm{A}_\alpha}^{a_\alpha} + k_{\mathrm{A}} C_{\mathrm{B}_1}^{b_1} \cdots C_{\mathrm{B}_\beta}^{b_\beta}. \tag{1.3}$$

In equilibrium state there holds $\frac{d\,C_{\mathrm{A}_1}}{d\,t} = 0$, etc., and one can define the constant characterizing the thermodynamic equilibrium:

$$K = \frac{k_{\mathrm{B}}}{k_{\mathrm{A}}} = \frac{C_{\mathrm{B}_1}^{b_1} \cdots C_{\mathrm{B}_\beta}^{b_\beta}}{C_{\mathrm{A}_1}^{a_1} \cdots C_{\mathrm{A}_\alpha}^{a_\alpha}}. \tag{1.4}$$

Linearity objective

One of the neocybernetic objectives is that of linearity. Clearly, the expression (1.4) is far from being linear — indeed, it is purely multiplicative. It turns out that applying a purely syntactic trick, linearity of the structures can be reached: Taking logarithms on both sides there holds

$$\log K' = b_1 \log C_{\mathrm{B}_1} + \cdots + b_\beta \log C_{\mathrm{B}_\beta} - a_1 \log C_{\mathrm{A}_1} + \cdots - a_\alpha \log C_{\mathrm{A}_\alpha}. \tag{1.5}$$

To get rid of constants and logarithms, it is also possible to differentiate the expression:

$$0 = b_1 \frac{\Delta C_{\mathrm{B}_1}}{\bar{C}_{\mathrm{B}_1}} + \cdots + b_\beta \frac{\Delta C_{\mathrm{B}_\beta}}{\bar{C}_{\mathrm{B}_\beta}} - a_1 \frac{\Delta C_{\mathrm{A}_1}}{\bar{C}_{\mathrm{A}_1}} + \cdots - a_\alpha \frac{\Delta C_{\mathrm{A}_\alpha}}{\bar{C}_{\mathrm{A}_\alpha}}, \tag{1.6}$$

where the variables $\delta C_i = \Delta C_i / \bar{C}_i$ are deviations from the nominal values, divided by those nominal values, meaning that it is *relative changes* that are of interest. The differentiated model is only locally applicable, valid in the vicinity of the nominal value.

Multivariate representation

A single reaction formula can also be expressed in a linear form when the variables are appropriately selected. However, to model complex systems consisting of various reactions, the data representation needs to be extended: The differing data vectors containing different sets of variables (the reactions employing different chemicals) have to be embedded in the same vector space to make them compatible.

Assume that the vector z is a vector containing all relevant variables capturing the state of the environment and the system itself, including, for example, relative changes in all chemical concentrations. This means that the vector Γ_i representing a single reaction can contain various zeros, assuming that the corresponding chemicals are not contributing in the reaction i. If the vectors Γ_i are collected as columns in the matrix Γ, one can write the individual expressions in (1.6) in the matrix form where one row is allocated to each of the reactions:

$$0 = \Gamma^T \, \delta z, \tag{1.7}$$

or, when written out,

$$\left\{ \begin{array}{ccl} 0 & = & \Gamma_{1,1}\delta z_1 + \cdots + \Gamma_{m,1}\delta z_m \\ & \vdots & \\ 0 & = & \Gamma_{1,n}\delta z_1 + \cdots + \Gamma_{m,n}\delta z_m, \end{array} \right. \tag{1.8}$$

where n is the total number of reactions, and m is the total number of chemicals. This expression needs to be compared to flux balance analysis: Now one only needs to study levels of concentrations, not changes in them. This is indeed essential in complex chemical systems, where the energy and matter flows cannot be exactly managed. The key point to observe here is that analysis of complicated reaction networks can be avoided: No matter what has caused the observed chemical levels, only the prevailing tensions in the system are of essence. The underlying assumption is that the system is robust and redundant: Individual pathways are of no special importance as there exist various alternative routes in the network.

It turns out that reactions can in principle be characterized applying linear algebra in the space of chemical concentrations. However, in practice it is not enough to only represent the concentrations if the properties of the whole system need to be captured. What else can the vector u contain?

1.3.2 Characterizing the metabolic state

The measurement vector z needs to be further studied to make it possible to capture all *internal tensions* in metabolic systems. As it turns out, the following extensions can, for example, be implemented without ruining the linear structure among the variables:

- **Temperature.** According to the Arrhenius formula, the reaction coefficients are functions of the temperature, reactions becoming faster as the temperature rises, so that $k \propto \exp(c/T)$. This means that when this is substituted in the formulas, and when logarithms and differentiations are carried out, the model remains linear if the new variable is defined as $z_T = \Delta T/\bar{T}^2$.

- **Acidity.** The pH value of a solution is defined in terms of a nonlinear formula: $\mathrm{pH} = -\lg C_{\mathrm{H}^+}$. Because it is essentially logarithm taken of a concentration variable, one can directly include the changes in the pH value among the variables, $z_{\mathrm{pH}} = \Delta\mathrm{pH}$.

- **Dissipation.** It has been assumed that the systems being studied are in thermodynamic balance. This homeostasis can be extended, however: The steady state can be determined not only in terms of the variables, but also in terms of their derivatives. This means that one can study *dissipative systems,* where the rate of change remains constant, a constant flow of chemical flowing into or out from the system. Looking at the formula (1.3), it is clear that model linearity is not lost if one has variables like $z_{\dot{C}} = \Delta\dot{C}/\bar{\dot{C}}$.

- **Physical phenomena.** It is evident that structures that are originally linear, like phenomena that represent diffusion between compartments, etc., can directly be integrated in the model, assuming that appropriate variables (deviations from the nominal state) are included among the variables.

In strong liquids one cannot always apply concentrations, but one has to employ *activities* instead, or actual activation probabilities. If it is assumed that these activities are some power functions of the concentration so that $\mathcal{A} = a_1 C^{a_2}$, after taking logarithms the model still remains linear in terms of the original concentrations. This means that — even though linearity is not compromised — the variables may become multiplied by some unknown factors, so that there is some scaling effect.

The vector z selected here is the measurement vector, containing *all* possible quantities that can affect the system behavior — internal system variables and external environmental variables alike. To have the actual *data vector* to be employed in modeling, the vectors first have to be preprocessed and appropriately scaled — these issues are studied in chapter 2. In practice, specially if the relationships between units are not clear, it can be motivated to carry out explicit data normalization to make data items better compatible (this issue is studied closer in chapter 3).

1.4 Case 2: Gene expression

However, the above studies are not the whole story — *genes* are an integral part of the metabolic system. Nature's way of implementing the genetic descriptions are mindbogglingly sophisticated — but when trying to capture the essence in those processes, perhaps one does not need to exactly stick to the nonidealities in the implementations? Here the goal of the system is seen as more important: The complicated mechanisms are only needed to reach the consistent balance among the system components.

1.4.1 Process of overwhelming complexity

Above, the domain of chemical reactions was studied and a coherent modeling framework was proposed for capturing the relevant variables that characterize the process state. However, the cell differs from other reaction vessels of organic chemistry: The genetic system is essentially a part of the metabolic system, controlling it. What can be said about it, how can this level be integrated in the system model?

The information required for building the proteins and controlling the cell metabolism is stored in deoxyribonucleic acid (DNA); the operational units in DNA are called genes. The proteins are synthesized in the process called gene expression: First the gene sequence is coded into messenger RNA in the transcription process, and, after being further modified and transferred from the nucleus into the cytoplasm, this code is compiled by ribosomes into proteins in the translation process. These proteins are either used as building blocks in the cell, or they act as enzymes, catalyzing other processes.

What makes this gene expression process specially complicated, is the fact that there exist feedback structures all the way along the process: First, the RNA molecules and proteins are "postprocessed" by various mechanisms controlled by some other genes and chemicals; second, some of the enzymes act as *transcription factors,* explicitly affecting the activity levels of other genes. In each case, there exist various complicated mechanisms (chromatin packing and dismounting, gene activation and inhibition, protein phosphorylation, myristylation, and glycosylation, etc.) how the interactions among the actors are implemented. It seems to be a hopeless task to accurately model the individual processes that are related in these processes (however, there exist various attempts to do that, for example [54]), and it seems that the system of interactions needs to be abstracted.

The gene interactions have been modeled applying abstract causality structures — an example of genetic networks is given, for instance, in [78]. Many different model structures have been proposed, for example *neural networks* [82]. However, there exist no unique gene regulation pathways, processes having very different time scales and relevances — simple projections onto a graph form can be misleading. Here, in the neocybernetic spirit, the pancausality idea is applied: In steady state, the transients have decayed, and all reaction chains have found their balance; the temporal sequences have changed to simultaneous patterns of effect flow, practically all genes participating in this equilibrium. How to characterize the internal tensions among genes?

1.4.2 "Cybernetizing" a genetic network

The genetic control system is much too complex to be modeled explicitly. The only possibility is to look at the genetic system directly from outside, studying the overall net effects. Again, when concentrating on the final thermodynamic balance among tensions, the time-domain complexities can be circumvented. There exist some clues.

Stationarity and statistics

Abstract away individual actions and realizations of interactions in the network, and assume that the stationary state has been reached. Is there anything one can say about such a system in general terms?

It has been observed that there exist peculiar similarities among very different kinds of complex systems. For example, it has been claimed (see [12]) and [5]) that distributions in self-organized complex networks statistically follow the *power law,* that is, there generally holds

$$z_j = c z_i^D \tag{1.9}$$

for some behaviors-related variables z_i and z_j, and constants c and D. Here, z_i stands for the free variable, and z_j is some emergent phenomenon related to the probability distribution of z_i. This power law dependency seems to govern all structures with fractal and self-organized structure For example, if z_i is the "ranking of an Internet page", and z_j represents "number of visits per time instant", the dependency between these variables follows the power law: There are some very popular pages, whereas there are huge numbers of seldom visited pages. As compared to Gaussian distribution, the power law distribution has "long tails"; the distribution does not decay so fast.

It is interesting to note that the power law distribution is closely related to another modern concept, namely *fractal dimension.* Assuming that the variable z_i represents some kind of "yardstick", determining the scale factor, and z_j represents the level of *self-similarity,* so that when one zooms the original pattern by the factor of $1/x_i$, there exist z_j copies of the original pattern (and this zooming process can be repeated infinitely), the fractal dimension of that pattern can be defined as

$$D = \frac{\log z_j}{\log z_i}. \tag{1.10}$$

When the pattern is simple, this definition coincides with the traditional ideas concerning dimension, but for complex patterns, non-integer dimensions can exist. Now, it is easy to see that, after taking logarithms, the parameter D in (1.9) closely corresponds to the fractal dimension for the networked system.

Multivariate nature and linearity pursuit

In the multivariate spirit, one can extend the single-variable formula (1.9) by including more variables — assume there exist μ of them:

$$1 = c' z_1^{D_1} \cdot \cdots \cdot z_\mu^{D_\mu}. \tag{1.11}$$

If there is only one variable z_i changing at a time, and one solves for z_j, this formula corresponds to (1.9). Furthermore, there can exist various such dependency structures — assume there are ν of them:

$$\begin{cases} 1 & = & c_1 z_1^{D_{11}} \cdot \cdots \cdot z_\mu^{D_{1\mu}} \\ & \vdots & \\ 1 & = & c_\nu z_1^{D_{\nu 1}} \cdot \cdots \cdot z_\mu^{D_{\nu\mu}}. \end{cases} \tag{1.12}$$

Now, if one takes logarithm on both sides of the formula, one has

$$\begin{cases} 0 & = & \log c_1 + D_{11} \log z_1 + \cdots + D_{1\mu} \log z_\mu \\ & \vdots & \\ 0 & = & \log c_\nu + D_{\nu 1} \log z_1 + \cdots + D_{\nu\mu} \log z_\mu. \end{cases} \tag{1.13}$$

It turns out that the multiplicative dependency has become globally linear — by only preprocessing the variables appropriately. To find a still simpler (locally applicable) model structure, one can further differentiate these equations around the nominal values $\bar{z}_i$, so that there holds

$$\left\{ \begin{array}{rcl} 0 & = & D_{11} \frac{\Delta z_1}{\bar{z}_1} + \cdots + D_{1\mu} \frac{\Delta z_\mu}{\bar{z}_\mu} \\ & \vdots & \\ 0 & = & D_{\nu 1} \frac{\Delta z_1}{\bar{z}_1} + \cdots + D_{\nu\mu} \frac{\Delta z_\mu}{\bar{z}_\mu}. \end{array} \right. \tag{1.14}$$

Again, the variables $\delta z_i = \Delta z_i / \bar{z}_i$ are the relative deviations from the nominal state. It turns out that there again holds the linear dependency, for variables having been preprocessed in an identical manner as in (1.7),

$$0 = \Gamma^T \, \delta z. \tag{1.15}$$

The scaling of the variables is not determined in a unique manner. As it turns out (in chapter 3) traditional normalization of the variable variances is motivated. Determining the "nominal state" is equally vague; if nothing else is known, it has to be assumed that extensive information of the system variables has been acquired, and the mean values characterize the nominal state. In what follows, the "Δ" symbols are dropped for brevity; however, it is still assumed that variations around the nominal state are small.

Combining the results of this section and the previous one, it can be claimed that *all phenomena that are relevant for characterizing the cellular state can be captured in a homogeneous linear framework* as shown in (1.15). It is interesting to study the properties of data properties that are dictated by the assumed structure.

1.5 Probability interpretations

When the above simple formulations were derived, individual phenomena were abstracted away. There are no more individual samples or time points visible, only their statistical long-term properties. Thus, it is interesting to briefly elaborate on probability distributions.

1.5.1 Fractality revisited

Study the outlook of the *multivariate fractal distribution.* Assume that one of the variables has been expressed in terms of the other variables

$$\log z_j = - \sum_{i \neq j} \frac{D_i}{D_j} \log z_i. \tag{1.16}$$

Variable $\log z_j$ is a weighted sum of assumedly large number of assumedly independent stochastic variables $\log x_i$. Because nothing more accurately about these variables is known, it can be assumed, according to the Central Limit

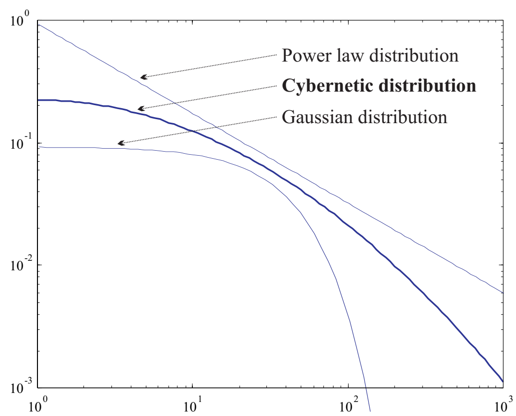


Figure 1.3: Schematic illustration of different distributions on the log/log scale

Theorem, that $\log y$ has normal distribution (or, indeed, *lognormal* distribution — see [55]):

$$p(\log z_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\left(\log z_j - \mu\right)^2/2\sigma^2\right). \tag{1.17}$$

Here, parameters c, μ, and σ are free parameters characterizing the outlook of the distribution. Taking logarithms, there holds

$$\log(p(\log z_j)) = c - \left(\log z_j - \mu\right)^2/2\sigma^2. \tag{1.18}$$

This means that the multivariate fractal distribution is *parabolic* rather than linear on the log/log axis, the three parameters being c, μ, and σ^2 (see Fig. 1.3).

Assume that the system variables can truly be characterized as having the dimension of *probability,* that is, genetic activity of a single gene can be seen as a probabilistic phenomenon. In such a case the above result gives new intuition. Indeed, the result is in conflict with “traditional” assumptions concerning fractal networks! This assumption seems to be supported also by evidence: For example, in Fig. 1.4, a manifestation of properties of a complex network are illustrated. It is clear that a parabolic curve better fits the observation points — the new model suits structures that are not strictly *scale-free.*

1.6 About more complicated distributions

In practice, model linearity cannot always be reached by as simple preprocessing of the variables as was presented above. For example, some genes can only be active in the vicinity of some location in the chemical data space — getting farther from that location in *any* direction makes the activity decay. The model structure can still be kept linear by appropriate selection and preprocessing of the variables; the key issue is to analyze how the system sees its environment.

If it is assumed that behaviors are results of high numbers of components interacting, the model is multiplicative with respect to the concentrations or probabilities. If logarithmic quantities are studied, one has additive models — one can assume that this assumption of log-linear behaviors can be extended beyond the

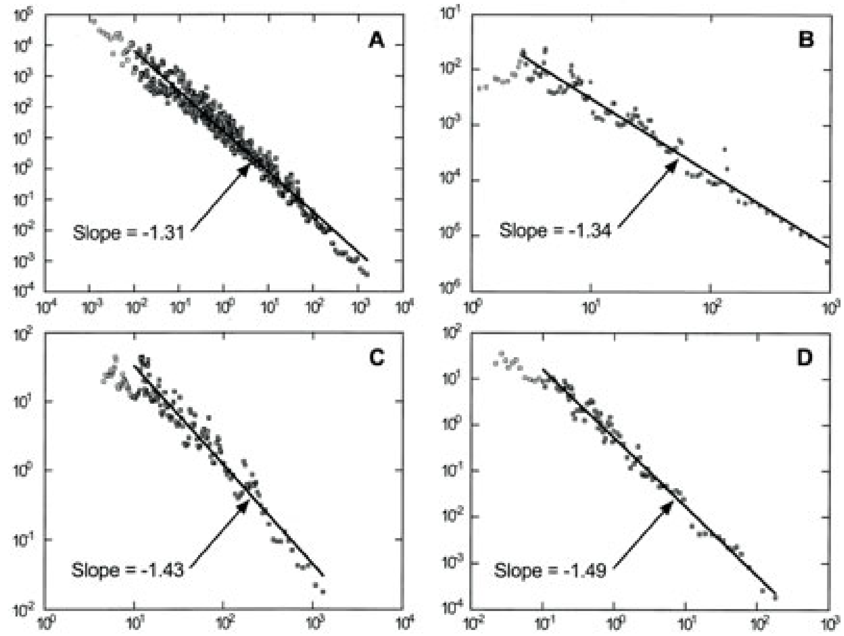


Figure 1.4: Properties of forest fires (from [52]): (A) 4284 fires on U.S. Fish and Wildlife Service lands (1986-1995), (B) 120 fires in the Western United States (1950-1960), (C) 164 fires in Alaskan boreal forests (1990-1991), and (D) 298 fires in Australia (1926-1991). The number of fires is given as a function of the burnt area

power law distributed variables. If the underlying distributions are Gaussian, one only needs to take into account the observations in the previous section: The behaviors can be assumed to be linearly related to *quadratic* functions of the input variables. This means that the set of input variables can (as the first approximation at least) be extended by including the squares of the most relevant variables, and products of them, among the input variables. Including the products of all variables among the features increases the size of the data space considerably.

This approach to representing general nonlinearities can be motivated in mathematical and in pragmatic terms. Mathematically speaking, smooth nonlinearities can be represented applying Taylor expansion, and such series can be approximated up to the second order when the quadratic terms are available. From the pragmatic point of view, the Gaussianity assumption is well in line with the Gaussian mixture model scheme (see chapter 6).

In this chapter, the homogeneity goal was reached: No matter what is the underlying realm like, the statistical properties of a complex network can be captured in a data vector, and it can be assumed that linear models are applicable if appropriate data preprocessing is applied. To truly capture all relevant variables, it is reasonable to include all variables that are potentially relevant — it is the task of the modeling machinery to select the most important of the variable candidates and to determine the dependency structures among them. Also, if there exist nonidealities giving raise to further nonlinearities in the system, the data vector can be augmented by the appropriate feature variables hopefully capturing the structure of nonlinearities. This means that the data vector can become very high dimensional, and the models to be studied explicitly need to be robust against high dimensionality. This is a real challenge — specially as tackling among the multitude of variables should be done not manually but automatically.

The first step towards a model of complexity, and towards deeper understanding of biological systems was taken in this chapter. However, the model structure (1.7) is barren, being very descriptive, and it is not suitable for real applications. Next, in the Level 2, the same model is first extended to capture the cell-scale phenomena, and after that, a more suitable formulation, or the structure of the "emergent models" is presented.

Level 2

Emergent Models of *Cellular Functions*

In the previous chapter, the data format was determined so that cell-specific information (or any network-originated information) could compactly be captured. The next task is to find the higher-level presentations, or the actual model structures, so that the underlying data can efficiently be exploited, and the essence of the cellular behaviors truly becomes manifested.

The key issue in this chapter are the *models,* or how to construct them in an appropriate way. It has to be recognized that *models are always false,* only showing a narrow projection of the complexity in real life systems. But good models can give intuitions.

Very simple mathematics only is employed here, and the model structures will be linear. There is nothing new in the mathematics — it is the interpretations that play the central role. Appropriate interpretations make it possible to escape from the reductionistic level to explicitly holistic models. These "emergent models" become practical when the components-oriented modeling view is exhausted. The new model structures can be seen as revealing the functions that take place in the complex system.

2.1 About "system semantics"

When searching for *good models,* philosophical questions cannot be avoided. It is such modeling issues that have been studied for millennia: What is the nature of systems, and how they should be represented. Indeed, what there is, what one can know about them — these problem fields are called *ontology* and *epistemology,* respectively (these issues are studied again in chapters 7 and 10). Here all these mutually related issues are collected under the common concept of *semantics:* What is the essence of a system, and how this essence should be interpreted?

Semantics conveys *meaning.* Traditionally, it is thought that semantics cannot exist outside human brain. However, to reach "smart models" that can adapt in new environments, one needs to make this meaning machine-readable and machine-understandable. Otherwise, no abstraction of relevant vs. irrelevant phenomena can be automatically carried out. Indeed, one is facing a huge challenge here, but something *can* be done.

For the purposes of concrete modeling, the notion of semantics has to be formalized in some way: This very abstract concept is given here very concrete contents, compromising between intuitions (what would be nice) and reality (what can be implemented in reality). It can even be said that a *good model formalizes the semantics of the domain field,* making it visible and compressing it. Now there are two levels of semantics to be captured:

1. **Low-level semantics.** The formless complexity of the underlying system has to be captured in concrete homogeneous data. The "atoms" of semantics constitute the connection between the numeric representations and the physical realm, so that the properties of the system are appropriately coded and made visible to the higher-level machineries. In concrete terms, one has to define "probes" and put them in the system appropriately.

2. **Higher-level semantics.** The high number of structureless low-level features have to be connected into *structures* of semantic atoms. Assuming that the semantic atoms are available, this higher-level task is *simpler,* being more generic, whereas finding representations for the low-level domain-area features is domain-area specific.

The former task — coding the domain-area information in concrete data structures — was studied in the previous chapter, whereas the latter task — connecting the atoms of information into relevant structures — is studied in this chapter.

The higher-level semantics determines how the information atoms are connected. In our numbers-based environments, a practical and robust approach towards capturing such *contextual semantics* is offered by correlations-based measures. If the data is defined appropriately so that it captures the dynamical balances in the system, the simple contextual dependency structures can also be seen to capture *cybernetic semantics* of the domain (see chapter 7). Assuming that information is conveyed in visible co-variations among data, structuring of lower-level data can be implemented by the mathematical machinery without need of outside expert guidance. Despite the trivial-sounding starting point, non-trivial results can be found when the mathematical structures cumulate. This makes it possible to reach "smart" models that adapt in unknown environments.

2.2 Constraints vs. degrees of freedom

The mathematical machinery has been traditionally used for solving engineering-like, reductionistic problems. However, the focus is changing: One should be capable of abstracting away the details and seeing the "big picture". In such cases one simply cannot go in the traditional bottom up direction — one has to go top-down, explicitly starting from the system level. And one cannot assume there is some existing *a priori* model structure — the models have to be based on observations. There are many challenges when new ways of thinking are exploited.

2.2.1 System models and identification

It is assumed that in a system the data are somehow bound together, and it is this bond that captures the essence of the system. The model structure derived in the previous chapter was of the following form, explicitly characterizing the bond between variables

$$0 = \Gamma^T z, \tag{2.1}$$

this matrix expression consisting of n separate scalar equations determining connections among variables in z. Indeed, this formulation is the very traditional approach to presenting structures within systems. For example, assuming that the matrix Γ consisting of a single column, and the data vectors $z(k)$, for k indexing the discrete time axis, are defined as

$$\Gamma = \begin{pmatrix} -1 \\ a_1 \\ \vdots \\ a_d \\ b_0 \\ b_1 \\ \vdots \\ b_d \end{pmatrix}, \quad \text{and} \quad z(k) = \begin{pmatrix} y(k) \\ y(k-1) \\ \vdots \\ y(k-d) \\ u(k) \\ u(k-1) \\ \vdots \\ u(k-d) \end{pmatrix}, \tag{2.2}$$

the connection among variables can be rewritten also in the form

$$y(k) = \sum_{i=1}^{d} a_i y(k-i) + \sum_{j=0}^{d} b_j u(k-j). \tag{2.3}$$

As it turns out, this is the traditional way of representing dynamics of a d'th order SISO (single input, single output) system. A huge body of theory has been developed, for example, for identifying the system parameters a_i and b_j based on a set of observations of the variables $y(\kappa)$ and $u(\kappa)$ for $k_0 \leq \kappa \leq k$ (for example, see [2]).

The models of the form (2.1) assume that the linear combination of the variables should be exactly zero — however, as the measurement values always are inaccurate, this does not exactly hold, and one has to extend the original model:

$$e = \Gamma^T z. \tag{2.4}$$

Here, e is the model error vector — the goal of identification of parameters in Γ is transformed into an optimization problem, where one tries to minimize the overall error variance. Very much effort has been put on enhancing the numerical properties of the identification algorithms, typically starting directly from the formulation (2.3), and for making them more reliable and robust — after all, the determination of the parameters is typically based on least-squares matching, and there are various reasons for problems [42].

First, a special challenge in traditional identification is caused by the nonideal noise properties. Different variables can be corrupted by the noise in different ways. And, in the case of colored noise, the uncorrelatedness assumption of the noise samples becomes compromised, and the parameter estimates become biased.

Second, if trying to capture all available information — by employing all available variables — in the models, as was proposed in the previous chapter, determination of the parameters sooner or later becomes an ill-defined task. As the large number of variables are more or less redundant, they are no more strictly linearly independent of each other, and the numerical properties of the algorithms can become very poor. For example, the variables $y(k-i)$ in (2.3) are in principle separate variables, but because of the smoothness in the signal behaviors, the variables are certainly not independent. The data covariance matrix (matrix that needs to be inverted in least-squares fitting) becomes badly conditioned.

In today's applications, these problems with high dimensionality severely plague the traditional modeling approaches. It is not only the high number of input data that causes problems, but the whole model structure is challenged. Applying the traditional model structure it is easy the implement SISO models, but one should also be capable of tackling with more complex systems consisting of various submodels — as in the metabolic system there exist various simultaneous balance reactions taking place at the same time. In principle, the data representation in (2.1) is naturally a MIMO structure, being a framework for presenting various simultaneous equations just as well. This structure only needs to be efficiently utilized. Are there alternatives to traditional ways of describing (locally linear) models?

2.2.2 Emergent models

The structure of the model (2.1) needs to be elaborated on: This can be accomplished as the model is interpreted in terms of linear algebra. Mathematically speaking, if there are μ separate variables, there are μ degrees of freedom in the data space, but each (linear) constraint decreases the number of degrees of freedom by one — specially, if there are ν linearly independent constraints, the number of remaining degrees of freedom is only $\mu - \nu$. The linear constraints constitute a *null space* within the data space: This means that in these directions there is no variability. The remaining $\mu - \nu$ directions in the data space constitute a linear subspace where all variation among variables is concentrated.

What do these degrees of freedom mean in practice? Originally, if there were completely separate unconnected variables (subsystems), there would be the maximum number of freedoms. When subsystems become connected, when interactions between them are established, the variables become coupled, thus reducing the number of free variables. Further, when feedbacks are introduced, the remaining inputs and outputs of the subsystems can still be connected. It

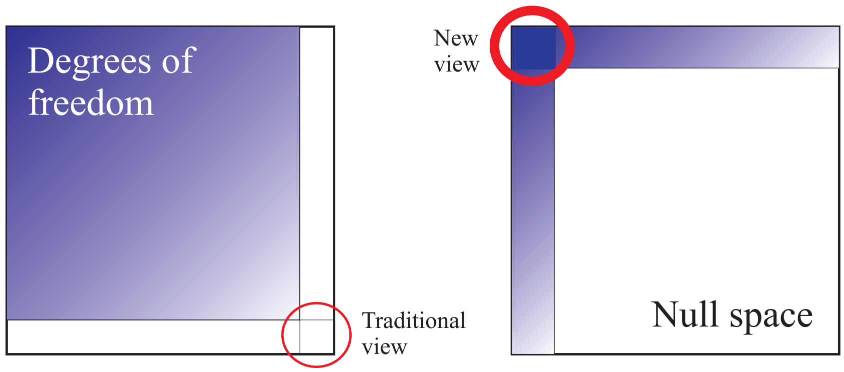


Figure 2.1: Schematic illustration of the covariance structure among data when there are few constraints (on the left), and when there are many constraints (on the right). The simplest presentation for the system properties changes as the number of constraints increases, or when the remaining degrees of freedom accordingly decrease

is specially typical in cybernetic systems where this scenario holds: Ability to recover after disturbances is a manifestation of tightly interconnected system. In such systems it is only a few degrees of freedom that remain more or less loosely controlled. In the metabolic system there are dozens of individual underlying reactions controlling the cellular metabolics, the chemical levels being balanced accordingly.

The key point here is that essentially the same dependencies among variables can be captured in terms of degrees of freedom as with constraints. At some point, when the number of constraints increases, the *most economical* representation changes: The simplest model with the least parameters is not the constraints-oriented model but the freedoms-oriented model (whatever it will be). According to the *Ockham's razor,* one needs to switch to *emergent models* when the system is cybernetic enough. In Fig. 2.1, the covariance structure of the data space is depicted: When the null space of constraints is dead and dull, all interesting behaviors are concentrated in the directions of remaining freedoms. It can be assumed that relevant phenomena in the cell are revealed by the "metabolic degrees of freedom"; it turns out that when applying very compact and behaviors-oriented models, the system starts looking more or less "clever" — indeed, speaking in such terms has to be postponed to next section.

Whereas the visible constraints are emergent patterns resulting from underlying dynamic attractors, the degrees of freedom make it possible to model this process of emergence and the structure of such patterns.

How is this dichotomy between constraints and freedoms manifested in concrete terms? For example, study an infinite-dimensional distributed parameter system that is governed by partial differential equations — a very natural way to characterize natural complex systems. As these PDE's are spatially discretized, there is a large number of ordinary differential equations connecting the local variables. Remember that only together with the boundary conditions the PDE's can uniquely determine system behaviors, thus giving rise to a very complicated system of hybrid equations that can seldom be solved explicitly. The constraints are now there in the form of dynamic and algebraic equations — the PDE's and the boundary conditions, respectively. The emergent behavior is typically manifested in terms of a low number of possible *modes.* For example, in the case of a vibrating plate, typically there only exist a few vibration modes; these "modes of freedom" can easier be modeled than the original constraints. Such freedoms-oriented approach is also quite natural, as then one directly concentrates on the time domain solutions of the equations that are immediately measurable in system behaviors.

Laws of nature are traditionally written in terms of constraints: The visible dependencies among observed phenomena are recorded. But, again, these surface patterns just emerge from underlying, more fundamental interactions. Perhaps one should rather start thinking in terms of "freedoms of nature".

It is difficult to escape the traditional ways of thinking: Traditional methods for analysis (modeling) and design (synthesis) are based on models that are based on constraints. And, indeed, constraints are the very basis of Wittgensteinian thinking: Languages are the means of structuring the world in terms of connections between concepts. This also holds what comes to formal languages like programming formalisms that the contemporary software tools are based on, and also traditional mathematics is based on finding ingenious proofs, or paths from a fact to another. Traditional mathematics exercises make nice pastime activity as the solutions typically are unique and hard to discover; however, the heavy mathematical machinery that is based on relevance is more general.

It is just as it is with detective stories: they make nice reading, but they are not plausible. Sherlock Holmes once said that "When you eliminate the impossible, whatever remains, however improbable, must be the truth". But in real life there are no clear-cut truths — modern detectives construct the "big picture" out from the mosaic of more or less contradictory evidence: The plausible explanation maximally fits the observations. This is today's world — as there is no unambiguous truth, it is *relevance* that is preferred; closer studies are needed here.

2.2.3 Towards inverse thinking

One needs to find appropriate mathematical formulations for the above intuitions. The leap is mainly conceptual — one has to go to the other end of the continuum, from structure orientation to data orientation. It is data originating from freedom structures that is more relevant than parameters originating from constraint structures. As it turns out, this approach makes it possible to avoid the age-old problem concerning symbolic and numeric representations: The structures are not fixed beforehand — or, actually, they are ignored altogether.

First, study the structure-oriented end of the continuum. For simplicity, assume that one wants to capture the nominal state when observations are available. Variations around the nominal state are interpreted — in the traditional spirit — as noise that should be eliminated from the model.

Assuming that there are many sources of noise, one can abstract away the properties of individual noise sources. According to the Central Limit Theorem, one can assume that the net effect of all noise sources is such that the error distribution is Gaussian, that is the observations are distributed along a high-dimensional bell-shaped curve around the mean value; it is this mean value vector ζ that is being searched for. For the data distribution one can write the density function

$$p(e) = \frac{1}{\sqrt{(2\pi)^{\dim\{z\}} \, |\mathrm{E}\{zz^T\}|}} \, \mathrm{e}^{-\frac{1}{2}\,(z-\zeta)^T \mathrm{E}\{zz^T\}^{-1}(z-\zeta)}. \tag{2.5}$$

In the spirit of maximum likelihood identification, one selects the best estimate for ζ by maximizing the overall probability of the measurements

$$\hat{\zeta} = \arg\max_{\zeta} \left\{ \mathrm{E}\left\{ p(e) \right\} \right\} \tag{2.6}$$

by adjusting the center of the distribution appropriately. Because logarithm is a monotonous function, maximization of (2.6) equals minimization of

$$-\ln p(e) = c + \frac{1}{2} \cdot (z-\zeta)^T \, \mathrm{E}\{zz^T\}^{-1} \, (z-\zeta) \, . \tag{2.7}$$

When looking at this goodness criterion, it is evident that the “natural” scaling of variables is reached if the measurements are preprocessed as

$$z' = \mathrm{E}\{zz^T\}^{-1/2} \, z. \tag{2.8}$$

In the space of these new variables z', it is simply the Euclidean distances (or their squares) that reflect the differences between vectors. This scaling explicitly emphasizes the null space directions where there exists no variation in the data space, thus boosting the constraints-oriented thinking. For the original data, the weighting matrix when evaluating distances is $W = \mathrm{E}\{zz^T\}^{-1}$; it is revealing to note that for Gaussian data this expression is called *Fisher information matrix.* Information is assumed to be in the inverses of covariances.

This is the today’s realm. The problem with the scaling (2.8) here is that if the dimension of z is excessive, the scaling matrix becomes badly conditioned: If there are linearly dependent variables, the inverse matrix cannot be found. In cybernetic systems the variables typically are highly redundant due to the high number of underlying constraints.

To proceed, one needs to look at (2.1): Even though the roles of Γ and z are intuitively clear, this can be incorrect intuition. Mathematically, if Γ is a vector, the roles of these two vectors are identical. There is duality among structure and variables: The visible manifestations of structure are numbers in vectors, just as the data is. It can be assumed that the information delivered by observations is distributed among the structure part and the data part. Normally, it is assumed that observations represent data — however, in this case when the constraints dominate, it can be assumed that *observations represent structure.* The situation needs to be turned upside-down: The information that is normally used for modeling is now regarded as noise, and only the “leftovers” not exploited by the traditional modeling approaches are concentrated on.

This kind of problems of traditional thinking can be concretized: For example, inverse covariance weighting results in excessive emphasis on linearly dependent variables, the identification procedures trying to distinguish between identical variables — what comes to representing the real properties of the data, such emphasis is counterproductive. Another example: When identification is carried out in the parameter space rather than in the data space, iterative adaptation steps trying to pull the parameters towards better locations, pathological effects can take place, specially, if the parameterization represents a dynamic model. The reason for this is that dynamic behaviors are related to the poles and zeros of the parameter polynomials rather than to the parameters themselves; convex combinations of parameter vectors do not necessarily reflect the properties of those vectors at all.

Now the model is constructed to capture the properties of the data directly, not the properties of some man-made parameterization.

What this intuition means in practice, what are the consequences? Traditionally, when searching for the structure, it is thought that variation outside the assumed structure is noise — now it is assumed that this remaining variation is interesting, reflecting those behaviors that have not been paralyzed by the constraints. Somewhat intuitively, one could employ the idea of *symmetry pursuit,* defining the data-oriented portion of the measurements as the inverse of the weighting in (2.8):

$$z'' = \mathrm{E}\{zz^T\}^{1/2}\, z. \tag{2.9}$$

This can be expressed also in another way: The symmetric weighting matrix among measurements becomes (see next section)

$$W = \mathrm{E}\{zz^T\}, \tag{2.10}$$

rather than being $\mathrm{E}\{zz^T\}^{-1}$, as in the (2.8) case. This means that directions of variation in the data are explicitly emphasized. What is nice is that no matrix inversions are needed, and such operations remain well-behaving even for high-dimensional data.

The motivation for the data weighting was here rather intuitive — however, in the next chapter this issue will be concentrated on from another perspective. It can be claimed that *such weighting mathematically corresponds to the view of data that locally controlled systems actually see* in their environments.

2.3 Technical exploitation

For the rest of this chapter, assume that the presented view of data were appropriate, and study the conceptual tools that are in place when this view is being functionalized. The approach to modeling here is synthetic rather than analytic: The approach is "technical", not trying to capture the actual underlying processes but only trying to imitate the results. The key point here is to present the best possible tools — multivariate statistical mathematics — and in the next chapter it is shown that there truly can exist some connection to real life.

2.3.1 Subspaces and mappings

It is beneficial to see the more general setting, or what the presented framework looks like when see from the point of view of mathematics and mathematical tools. When functionalizing the freedoms-based model structure, one faces a pattern matching problem where linear algebra is needed.

Data preprocessing

Forget about all connotations that the variables in z may have, and apply conditioning to this data so that the technical assumptions become optimally fulfilled.

The first assumption is that of model linearity. Typically, problems are caused by the fact that data from linearized models are *affine,* that is, additive constants are needed in formulas. To get rid of the affine terms, the data can be transformed to follow a strictly linear model, for example, applying mean-centering — this is the standard approach when doing strictly data-based modeling where the nominal values of the variables are not known.

However, these problems are only faced when doing constraints-oriented modeling: When concentrating on the freedoms, no mean-centering is necessary.

The second assumption is quadratic nature of cost criteria. The reason for this is that easily manipulated and explicit formulas can be reached. The quadratic criteria mean that variations in the data are emphasized, and to reach reasonable models, appropriate scaling of data needs to be carried out. Assuming that all variables are equally informative, different variables can be equally "visible" by normalizing them to have unit variance, because units are arbitrary. This means that one uses either correlation matrices (if data is mean-centered) or cosine matrices (if data is not centered) as association matrices (see [92]).

Whatever are the data preprocessing steps, the original data z will hereafter be denoted ζ.

The data scaling is very crucial, affecting the results very much – the normalization should be motivated better. Indeed, as shown in Sec. 3.3, if the data is coming from a truly cybernetic system, it turns out that normalization is the *natural* way of seeing inter-system signals.

Pattern matching

In concrete terms, the freedoms-based model characterizes the location of an observation in the data space in terms of the degrees of freedom. The degrees of freedom are manifested as n linear *feature vectors* φ_i being collected in the matrix φ. Because of linearity, features can be freely scaled and added together. The observed patterns, combinations of the variables, are assumed to be weighed sums of such features, so that one can write

$$\hat{\zeta} = \sum_{i=1}^{n} \xi_i \varphi_i = \varphi \xi, \tag{2.11}$$

where ξ is the vector of weighting factors. If vectors φ_i are seen as coordinate axes, ξ_i are the coordinate values. Use of the feature model becomes an associative pattern matching process against data.

Assuming that $n < m$, arbitrary variable combinations ζ cannot be exactly represented by the features, and when searching for the best possible match, or estimate $\hat{\zeta}$, one is facing an optimization problem. When the representation error $\zeta - \varphi\xi$, weighted appropriately, is minimized, one can write the quadratic criterion

$$J(\xi,\zeta) = \frac{1}{2}\left(\zeta - \varphi\xi\right)^T W \left(\zeta - \varphi\xi\right). \tag{2.12}$$

The unique minimum is found when the gradient vector is set to zero:

$$\frac{d\,J(\xi,\zeta)}{d\xi} = \varphi^T W \varphi\xi - \varphi^T W \zeta = 0, \tag{2.13}$$

giving the unique solution

$$\xi = \left(\varphi^T W \varphi\right)^{-1} \varphi^T W\,\zeta. \tag{2.14}$$

In practice, this implements a mapping from an m dimensional space of ζ onto the n dimensional subspace of ξ spanned by the feature axes. Variables in ξ are called *latent* or *hidden variables.* Because of the data compression, exact match is not found, and one can only hope that the ignored variation is *noise,* not actual *information.*

How to distinguish between noise and information, then? Formally, there is no difference in the manifestation of variations in the data, and one has to apply *ontological assumptions* concerning the nature of relevant properties in the data. First, following the above discussions, one should select the weighting matrix as

$$W = \mathrm{E}\{\zeta\zeta^T\}. \tag{2.15}$$

Selection of the feature vectors so that they would represent the most important degrees of freedom can also be explicitly solved, and the solution is given by PCA presented in Sec. 2.3.2. This means that one should choose φ so that the *subspace of the n most significant principal components of data* is spanned by columns φ. Indeed, it is not necessary that the features are exactly the covariance matrix eigenvalues, $\varphi_i = \theta_i$, but it suffices that there holds

$$\varphi = \theta D \tag{2.16}$$

for some orthogonal transformation matrix D, so that $D^T = D^{-1}$. Correspondingly, the latent variables are modified as $\xi' = D^{-1}\xi$. The optimal selection of features is also non-unique — regardless of how the (non-singular) basis is constructed out from the matrix θ, the same variation can be captured. This means that after PCA, different kinds of *factor analysis* techniques, rotations, etc., can be applied to find a physically better motivated basis. For example, if the variables ζ have some constant bias, so that they are not zero-mean, it is possible to determine variables ξ that also have non-zero-mean — they can even always remain positive. When the variables represent some physical quantities, such non-negative coding is more plausible.

Regression based on latent variables

When the latent variables ξ are available, they can be exploited, for example, for *regression,* mapping data from the latent basis ξ onto some output space of y, so that $y = f^T\xi$. If the mapping is implemented through the low-dimensional latent basis rather than directly from the variables ζ, noise gets filtered out, and more robust estimates for the output can be found.

Similarly as above, the criterion for a good mapping model is minimization of quadratic criterion. When written for a single output variable y_j at a time, the mapping error becomes $\epsilon_j = y_j - f_j^T\xi$, and a reasonable criterion is found when the variance of this error, or $\mathrm{E}\{\epsilon_j^2\}$, is minimized. However, in some badly conditioned cases a generalization is in place: Robust regression models are found when the *regularized* criterion is applied where the parameter sizes are also emphasized:

$$J_j(f_j) = \mathrm{E}\{\left(y_j - f_j^T\xi\right)\left(y_j - f_j^T\xi\right)^T\} + \frac{1}{q} f_j^T f_j. \tag{2.17}$$

When the gradient is set to zero,

$$\frac{d\,J_j(f_j)}{df_j} = 2\left(\mathrm{E}\{\xi\xi^T\} + \frac{1}{q}I_n\right) f_j - 2\mathrm{E}\{y_j\xi^T\}^T = 0, \tag{2.18}$$

one can find the unique solution:

$$f_j = \left(\mathrm{E}\{\xi\xi^T\} + \frac{1}{q}I_n\right)^{-1} \mathrm{E}\{y_j\xi^T\}^T. \tag{2.19}$$

When this procedure is carried out for all outputs y_j separately, one can see that essentially the same formula is found in each case, and one can write a combined expression for all individual mappings as

$$f = \left(\mathrm{E}\{\xi\xi^T\} + \frac{1}{q}I_n\right)^{-1} \mathrm{E}\{y\xi^T\}^T. \tag{2.20}$$

If there is no need for regularization, that is, if the covariance $\mathrm{E}\{\xi\xi^T\}$ is invertible, one can use the standard formulation

$$f = \mathrm{E}\{\xi\xi^T\}^{-1}\mathrm{E}\{y\xi^T\}^T. \tag{2.21}$$

A special regression case is where the output is chosen to be the original data, $y = \zeta$, so that *reconstruction* of the data is being carried out, noise hopefully being filtered out during the compression process:

$$\hat{\zeta} = \mathrm{E}\{\zeta\xi^T\}\mathrm{E}\{\xi\xi^T\}^{-1}\,\xi. \tag{2.22}$$

Now there are technical tools for implementing mappings from data onto the feature subspace and back. The remaining problem is the determination of that feature subspace.

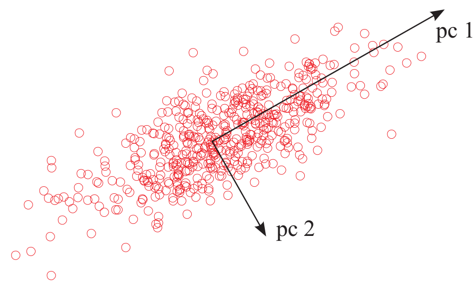


Figure 2.2: Principal component analysis reveals the variation structure in data

2.3.2 Multivariate tools

The distinction between constraints and freedoms can be elaborated yet in another way. Remember that traditionally one wants to minimize the sum of squared errors over the set of measurement data:

$$\Gamma = \arg\min_{\Gamma} \left\{ \mathrm{E}\{e^T e\} \right\}, \quad \text{when} \quad |\Gamma_i| = 1 \quad \text{for all} \quad i. \tag{2.23}$$

In this vector formulation, to have a well-conditioned optimization task, one has to fix the model vector size to avoid trivial solutions $\Gamma_i = 0$ (this is reached by introducing the additional restriction $|\Gamma_i| = 1$). This constrained optimization problem results in search for constraints in the traditional sense — indeed, the solution here is the method called *Total Least Squares.* When searching for the freedoms instead, the objective is exactly opposite:

$$\varphi = \arg\max_{\varphi} \left\{ \mathrm{E}\{\xi^T \xi\} \right\}, \quad \text{when} \quad |\varphi_i| = 1 \quad \text{for all} \quad i. \tag{2.24}$$

Note that even though it is freedoms that are searched, the mathematical machinery again is based on constrained optimization — constraints simply are the kernel of today's models! Here, vectors φ and ξ have been used to emphasize their different roles as compared to Γ and e: Defining $\xi = \varphi^T \zeta$, it is now the "error" ξ that is to be maximized, and φ is the axis along which this maximum variation in data is reached. If the vectors Γ_i and φ_i are interpreted as directions in the data space, mathematically speaking they reveal maximum orthogonality and maximum parallelity among these vectors and data, respectively. Applying the objective (2.24), it is assumed that variation in data is interpreted as information, whereas traditionally variation is seen as noise. And, specially, it is *covariation* among variables that carries information: Covariations can reveal the underlying "common causes" that are reflected in the measurements.

The solution to the problem (2.24) is given by *principal component analysis* or *PCA* (for example, see [6]). Without going into details (for example, see [42]), the basic results can be summarized as follows.

> The degrees of freedom can be analyzed using the data covariance matrix $\mathrm{E}\{\zeta\zeta^T\}$. The variability is distributed in the data space along the eigenvector directions of this matrix, variance in the eigenvector direction θ_i being given by the eigenvalue λ_i:
>
> $$\mathrm{E}\{\zeta\zeta^T\}\theta_i = \lambda_i\theta_i. \tag{2.25}$$

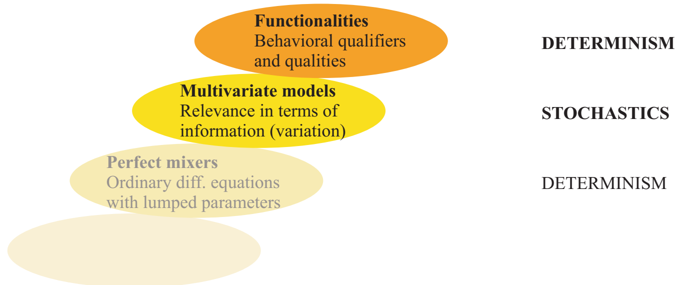


Figure 2.3: The remaining levels in the hierarchy of models in Fig. 3

Principal component analysis gives a structured view of freedoms in the data: The axes θ_i corresponding to most significant eigenvalues span a subspace where most of the variation in data is found (see Fig. 2.2). If $n < m$, meaning that the high-dimensional data is projected onto a lower-dimensional principal subspace, data compression takes place where the data variation is maximally preserved: If an n dimensional PCA basis is exploited, the model captures $\sum_{i=1}^{n} \lambda_i$ of the total variation in data — assuming data normalization, this total variation is $\sum_{j=1}^{m} \lambda_j = m$.

It turns out (because of symmetricity of the matrix $\mathrm{E}\{\zeta\zeta^T\}$) that the eigenvectors are orthogonal (indeed, orthonormal), so that the principal component directions can be used as a well-conditioned subspace basis vectors in a mathematically efficient way.

Principal components offer a very powerful mathematical framework — but is it physically meaningful? Complexity intuition says that self-organization of structures necessitates some kind of nonlinearity and instability: To reach emergence of differences, one needs positive Lyapunov exponents in functions, and to stabilize such divergent processes, nonlinearity is needed. However, as analysis of PCA reveals, there exist structures in data that can be motivated also in linear terms and using stable dynamic characterizations. Indeed, as will turn out later in chapter 3, the PCA intuitions will be of crucial importance when studying the properties of cybernetic systems. This means that the emergent patterns are very different as compared to the traditional chaotic images; the PCA patterns are based on global rather than local properties of functions.

2.3.3 New levels in emergence hierarchies

In Fig. 3, it was shown how deterministic and stochastic approaches can be seen to alternate in the hierarchy of emergent levels. Now the multivariate statistical models determine yet another stochastic level above the highest deterministic one: Information from the lower levels is extracted in the form of variations, and among that data, statistical dependency structures are determined in terms of covariations (see Fig. 2.3). Because such covariation structures can be found applying convergent algorithms, one is escaping the (mental) deadlock: Structures *can* emerge even in balance systems, one does not always need chaos and positive feedback to shake the underlying structures apart.

But a further transition from this stochastic level to a yet higher deterministic level is more or less straightforward. If there are statistical structures that can be employed to compress the statistical data, such abstracted phenomena can be named, thus introducing new distinct concepts. As the statistical structures represent dynamical balances, the essence of such concepts is that they are *attractors* in the data space dictated by the properties of the environment. The domain-oriented "concepts" are manifested as *emergent functionalities.* These functionalities are non-programmed, they are not explicitly designed; they do not reflect the intentions of humans but the properties of the interplay between the system and the environment.

Using such higher-level concepts, the functioning of a complex system can be appropriately structured: there are names for behaviors that are assumedly relevant, being manifested in observations. A "natural language" based on such concepts would be beneficial when trying to characterize and understand the functioning of a complex system. Based on the low-level semantics, interpretations of the emergent concepts are self-explanatory.

However, complex systems differ from each other, and the "axes of relevance" cannot always be defined in such a straightforward way. For example, in industrial plants it is the *quality measures* that are the most important quantities when characterizing the plant operation. The industrial plants do not simply reflect their environment; they are constructed for some special purpose, and the qualities cannot be dictated only by the environment, but the intentions of the system designers have to be taken into account. Generally in technical systems the operation (and "evolution"; see chapter 3) is goal-directed — rather than reflecting the environment directly, the emergent structures should reflect the coupling between the input space (environment) and the output space (qualities). Rather than employing PCA, the model structures should implement the cross-correlations among the two spaces. For engineering-like development of the processes, or "artificial evolution", there are other regression techniques available, for example PLS and CCR (see [92]).

Natural systems are simpler than technical — assumedly they just want to survive, trying to match with their environments (see next chapter), so that one can employ the PCA-based models directly. One can assume that it can only be the visible variables that determine the observable behaviors; if the variables are selected and scaled appropriately, there is no reason why a mathematical machinery could not capture the same phenomena that are followed by the biological machinery. A more detailed example of emergent-level modeling is given below. The emergent functionalities reflect match with environment; as seen from above, such behavior seems *clever* in its environment.

2.4 Towards system biology

Finally, study how the presented approaches can be exploited for modeling cellular systems in practice — and how they perhaps could be exploited.

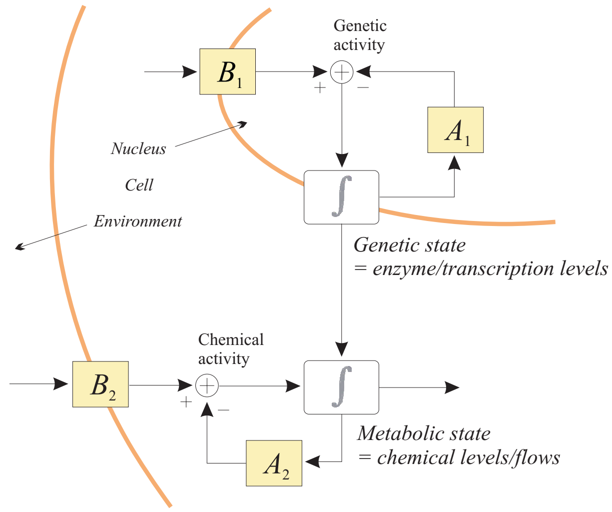


Figure 2.4: Two time scales in the cellular system

2.4.1 Facing real systems

Traditionally, modeling of biological (cellular) systems has been carried out in the rather traditional spirit: The goal has been to determine the constraints, individual dependency structures, exploiting more or less straightforward, SISO-type model structures. Data collection of cellular metabolics is typically carried out applying one-variable-at-a-time experiments. Similarly at the genetic level: Single "knock-out genes" can be explicitly deactivated to study their functions, resulting in non-natural behaviors. One reason for these simplified approaches are the practical problems, as the cellular state is difficult to measure, but new solutions are being introduced (for example, the microarray techniques for measuring the whole array of genetic activities simultaneously). Indeed, today there exists plenty of data, but this data is not necessarily well-conditioned or of good quality: The level of measurement noise is high.

Reaching reliable measurements is challenging, because the responses vary in different circumstances when the environment changes. What is more, because of buffering effects in the cells, huge dosages of reagents are needed (single input) to have noticeable responses (single output). On the other hand, these effects cannot be focused, being reflected to the whole set of variables. The experiments do not really characterize typical behaviors — the cell may become crippled altogether. Another traditional problem in metabolic systems is that they seem to be highly redundant (this also applies to gene expression). It seems that there typically is not just a single reaction mechanism explaining the processes, making it difficult to uniquely identify causal structures and model parameters. What makes this still more difficult is that not all chemicals can be recorded, and not all reactions are even known.

All of this suggests that multivariate statistical methods are needed. When applying the multivariate methods, buffering is just a manifestation of the internal feedbacks, and observations of the new balance deliver valuable information concerning the underlying metabolic processes and functions. No one-input/one-output studies are needed. Also the problems with unclear causal dependencies are avoided because of the pancausality assumption: First, the actual reactions are not searched for, but the "residual" variations; second, PCA is just the right tool to model redundant and noisy phenomena, because it transforms from the visible variables to new latent variables, where noise and redundancies among variables has been ripped off. Putting it freely: "If they are there, but if they cannot be seen, just ignore them". All relevant variables and dependencies cannot be detected, but they can be ignored as long as they do their job in maintaining the system balance.

As studied in chapter 3, the PCA-based model structures are motivated not only from the data analysis point of view. It can be claimed that in evolution it is the principal subspace that is naturally being pursued by surviving systems that are capable of most efficiently exploiting the environmental resources.

The objective here is to study living cell rather than pathological cases. Balances are more characteristic than transients, and it is steady states that are modeled. Because metabolic processes are well buffered, remaining near the nominal state, linear models are locally applicable. Rather than carrying out tests in a SISO manner, the whole grid of chemicals are studied simultaneously. This applies also to the transcription factors on the genetic level: As studied in chapter 2, genetic networks can be modeled applying the same model structures as the chemical processes — the metabolic processes are fast, whereas the genetic ones are slow (see Fig 2.4). In the figure, the linear pattern recognition processes are expressed in terms of dynamic state-space models, implementing two overlapping processes levels of "generalized diffusion".

Both of the levels can be combined in one model structure, and all information can be included in the data vector. The modeling procedure goes as follows: The sets of metabolites, transcription factors, and relevant environmental conditions (temperature, pH, ...) are defined, and experiments are carried out in different conditions, collecting data during the transients and in steady state. The degrees of freedom are found, determining the metabolic and genetic functions. Data orientation is necessary, multivariate tools are needed as the signal details are abstracted away, whereas emergent long-term phenomena become visible. Stationarity and validity of statistical measures is assumed — however, this assumption does not strictly hold. When the system becomes more and more complex, and as the number of constraints increases, the situation becomes blurred: some of the constraints are more acute than the others, and the thermodynamic balances are not necessarily all reached instantly.

2.4.2 Case example: Modeling genetic networks and metabolic systems[1]

In the project SyMbolic (Systemic Models for Metabolic Dynamics and Gene Expression), funded by National Technology Agency of Finland (TEKES) during 2004 – 2006, new kinds of models were derived for representing the cellular dynamics, and one of the approaches was the exploitation of the idea of emergent models.

As an application example, modeling of data from yeast cell cultivations were

[1] The simulations were carried out by Mr. Olli Haavisto, M.Sc.

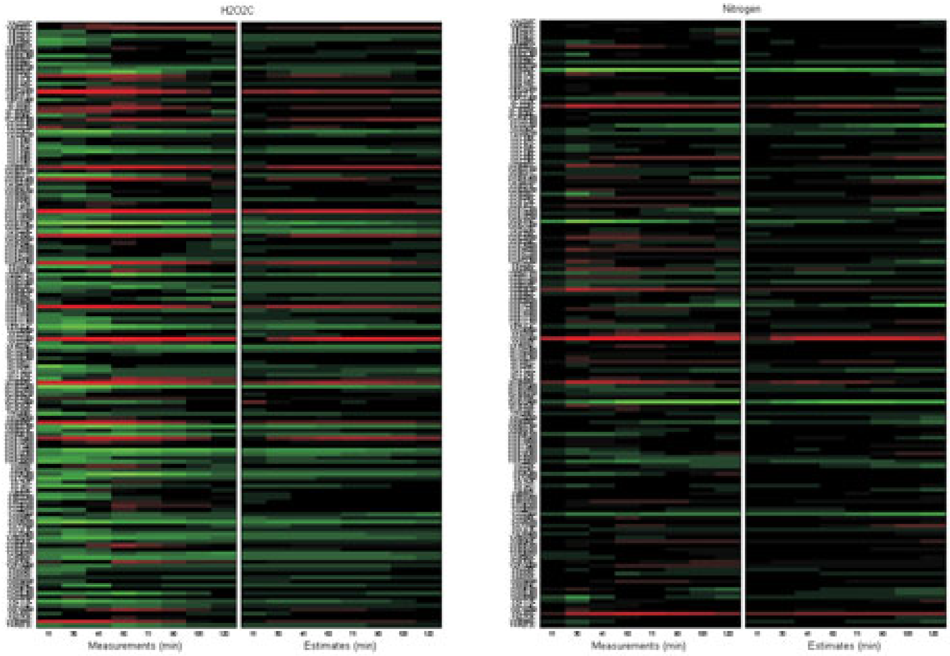


Figure 2.5: Two open-loop experiments with the model, showing 256 "stress genes" (red color meaning activity increase, green meaning activity decrease). Horizontal axis is time, and the rows represent individual genes. In the leftmost figures, hydrogen peroxide step is being simulated for two hours, and in the rightmost ones, nitrogen step is simulated. In both cases, the actual behaviors in the genetic state are shown on the left, and the estimates given by the four-state model are shown on the right. Despite the transients, there is a good correspondence between the observations and the very low-dimensional model (see [35])

used (see [35]). There were a few dozen experiments available (from [15] and [32]), where different kinds of step changes in the environment had been executed, and the resulting gene activity transients had been recorded. The step experiments were interpreted to present "stress responses" of the yeast cells. Modeling this data was quite a challenge, as there was not enough data, and not all data was quite reliable. Indeed, there do not exist many reports of dynamic modeling of the cellular behaviors (one attempt that is also based on latent variables can be found, for example, in [39]).

Because the available data was in the form of step experiments, the model was restructured so that the experimental setting was captured: The causal structure from manipulated variables to observations was simulated in the model. The environmental variables (substrate properties, temperature, etc.) were collected in the input vector u, and the gene expression levels were collected in the output vector y. Rather than constructing a traditional static PCA model between these data sets, a dynamic model was constructed applying so called *stochastic-deterministic subspace identification* (see [60]). This means that also the time sequence among data is taken into account and exploited when the latent variables x are constructed, the subspace identification algorithm automatically constructing a discrete-time state-space model (see [4]) for "generalized diffusion". Such a model can be efficiently exploited for implementing, for example, *Kalman filters* for optimally estimating the system state (see [24]). It can also be claimed that there is a connection to *Hidden Markov Models* here: The state sequences are reconstructed optimally, even though the probability interpretations are violated (this interpretation becomes more appropriate when the state variables are kept strictly non-negative; see chapter 6).

The dimensions of the vectors were selected so that m was about ten, and there were about 4000 output variables; the number of latent variables n was chosen to be 4.

When there are explicit transients in the data, the underlying assumptions about system stationarity are violated. This gives raise to model errors: There are slower and faster reactions taking place, some reaching their balance faster than the others. Indeed, a "Pandora's box" is opened when the balance assumption is abandoned — "extra" behaviors become visible in stress (transient) situations. What is more, complex transient reactions can take place in parts, where subprocesses follow each other; each of such intermediate products spans a new dimension in the variable space, and each chemical reaction introduces a new constraint, compensating for the increased dimensionality only after the balance is reached. The net effect is that the *invisible* dimensions in the variable space become visible during changes.

The assumption beyond the adopted modeling approach is, however, that balances are more characteristic to cellular systems than the transients are. And, indeed, it seems that at least the steady states are nicely modeled, whereas the transient behaviors are not reproduced as well by the model (see Fig. 2.5). Still, it seems that the extreme compression of the variable space does not ruin the steady-state correspondence. Truly, there seem to exist only few degrees of freedom left in the behavioral data.

2.4.3 "Artificial cells"?

When the presented model structure is seen in a perspective, it seems to open up new horizons. Using some imagination, it is easy to draw interesting interpretations.

It can be claimed that the degrees of freedom in a cellular system characterize *metabolic behaviors* or *functions.* When the environment changes, the new balance is found along these axes in the chemical space when "chemical pattern matching" is carried out. For example, assuming that available glucose goes up, it is also mannose production that goes up, or some other processes that exploit glucose. In fact, there is only balance pursuit taking place: But after "anthropocentric", finalistically-loaded interpretations are employed, when some chemicals are interpreted as nutrients, some others as metabolic products, and the rest as waste, one reaches "emergent interpretations". When complexity cumulates, the balance reactions start looking goal-oriented, pre-planned, and "clever". Scarcity of some chemicals changes the balance appropriately, trying to compensate for the shortage, and abundance results in the opposite outcome, as being visible in the "activity vector" ξ (or x).

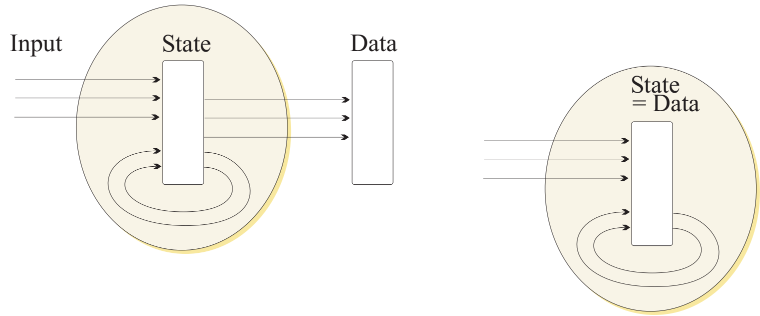


Figure 2.6: From *data modeling* (on the left) towards *system modeling* (on the right). The variables being measured are system variables (because of pancausality, changing them also changes the system state), and model structures being exploited are those of the system itself (both being based on the principal components of the measurement data; see next chapter)

A system model can be applied also for design and control. When the variables are selected appropriately, so that system semantics is captured, and if the pancausality assumption holds, the constructed modes are not only *data models* — they are *system models*. They can capture the fundamental essence of systems and system-specific variables. They can be used not only for monitoring, but also for design and control construction: Changing variables appropriately also changes the resulting balance (see Fig. 2.6). The remaining degrees of freedom in the system reveal the possibilities of further controls to make the system still more balanced; in this sense, *process data mining* or real *knowledge mining* becomes possible, where information can be gathered directly from the behaviors, not from model-based assumptions. New kinds of models make it possible to implement new kinds of controls — higher-level controls. However, new challenges are faced: When new feedbacks are introduced, the set of freedoms changes. Control design becomes an iterative task, and new kinds of design tools are needed.

The ideas of biological cybernetic systems can be extended to technical (bio)processes: The still unbounded degrees of freedom can be regulated, new feedbacks can be constructed, so that still better balanced higher-level "superorganisms" are constructed. On the other hand, the "broken" control loops can be fixed in the same way: For example, if the glucose level varies in the body more than it should, this can be compensated by insulin injections — along these lines, diabetes is treated manually today; but a simple automatic control loop could be implemented also as a step towards better lives of the "cybernetized patients".

Today, there are problems when trying to implement such integrated systems. For example, the glucose sensors need regeneration after a short time; after this problem is solved, new ones are sure to emerge. The key challenge is not how a single functionality — like sensitivity to a certain chemical — could be implemented, but how to keep the new system in a sustainable balance with its environment. This goal sounds very cybernetic. Indeed, it is the whole engineering-like thinking that has to be abandoned: Whereas one today concentrates only on a single functionality, it is the whole entity that has to survive in the complex environment. The same challenges are faced in all applications of tomorrows medicine: If the new integrated systems are not in balance, the body rejects the transplants. Finding such balanced solutions is a holistic problem that cannot be solved reductionistically. One needs to change the whole way of thinking from invasive to humble: One has to admit that nature's own structures offer the most useful adapted solutions to the key problem, that of finding a sustainable equilibrium in the metabolic system. Indeed, there exist ready-to-exploit cell structures to be used as platforms for new functionalities; one only has to take the next step and tame and cultivate the bacteria, domesticating them. Rather than constructing completely new artificial cells, one has to obey those ways of thinking that nature has followed: New structures are constructed on existing ones, just redirecting and boosting the evolution.

This all does not only apply to medical engineering: The key challenge in future industrial systems is their life-long maintenance. It would be reasonable to implement some level of cybernetic self-repair or adaptability in those systems, too, rather than only fixing the broken parts. Tomorrow's industrial systems also need to be in balance with their surroundings, not fight against it.

The presented emergent models were just models, and models should not be mixed with reality. For example, how could one motivate the "chemical pattern matching" as a fundamental cellular principle? How could a system with no central control accomplish it, even if it would like to do it? And, to reach *real system biology,* it is not only the internal behaviors within the cell that need to be captured — the next level is the coordination among the cells, and, generally, among populations. The challenge is to find out how such orchestration can be explained in terms of local actions only.

When studying natural systems, it is difficult to get farther only studying available data and existing systems — one needs stronger modeling principles. One should not only try to explain phenomena: One should proactively try to find the underlying principles. This kind of ideas crystallize in the question: *What are the goals of systems?*

Level 3

Elasticity of Systems and *Goals of Evolution*

As it was observed in the previous chapter, the biological data can efficiently be modeled applying multivariate statistical tools. However, it seems that such data-oriented approaches do not suffice after the elementary levels. It is the same with "chemical pattern recognition" as it is also in other areas of data-based modeling: The statistical correlations are not enough to unambiguously determine the higher-level structures.

To get further, one has to apply more ambitious ways of limiting the available complexity. Traditionally, the approach is to introduce more stringent and complex model structures to direct the parameter matching. However, again following the neocybernetic ideas, no extra complexity is voluntarily integrated in the models — alternatives to increased complexity are searched for instead.

The alternative employed here is rather radical.

In the postmodern era, there should exist no taboos. However, one thing that has never been proposed in circles of modern serious science, is that of *finalism.* One should only answer the *how* questions, never the *why* questions. Yet, applying teleological assumptions, most compact problem settings are reached, and one can also study systems that *not yet exist.* The claim here is that appropriate finalistic arguments can also be given concrete contents, so that they become verifiable — or falsifiable. It is only the starting point that sounds radical: The discussions collapse back into well-established frameworks.

It has been said that nothing in biology can be explained without taking evolution into account. And here this observation is exploited by studying the question: What is it that evolution tries to accomplish? Such issues are studied in the neocybernetic perspective — balance pursuit is the only finalistic goal after all, together with extreme environment-orientedness.

3.1 Balancing between static and dynamic models

From now on, one needs to (for a moment) forget about the technically oriented derivations in the previous chapter. The emphasis is changed: It is not what the model designer intends that is relevant — the interesting things are those what the system naturally does. Again, there is the same starting point (1.7) that is assumed to hold for a thermodynamically consistent chemical balance system. Indeed, such a simple formulation can be written for any linear system, no matter what is the domain field, if the variables are selected appropriately. This set of equations can be interpreted so that it defines a *static balance* with *no structure,* and one first needs to extend the framework.

3.1.1 Restructuring data

Assume that the variables in z in (1.7) are divided in two parts: Vector u, dimension m, describes the environmental conditions, whereas vector $\bar{x}$, dimension n, contains the system-specific *internal variables,* somehow characterizing the equilibrium state of the system. The internal state is not assumed to be necessarily observable by an external observer. The "environment" here is not something external — it only consists of variables that are determined from outside, but essentially all variables (concentrations) still coexist in the same volume. Rewriting the constraints characterizing the system, one can distinguish between the variables:

$$A\bar{x} = Bu. \tag{3.1}$$

The construction of the matrices A and B is not uniquely determined by this expression — this issue, determination of the system matrices in a plausible way, is studied later. To keep the internal state of the system well-defined, it is assumed that there are as many constraints here as there are latent variables, so that A is square. Because of environment-orientedness, the internal variables are assumed to be directly determined by the environment, so that there assumedly is a (linear) dependency between $\bar{x}$ and u. Formula (3.1) is an implicit expression; assuming that A is invertible, one can explicitly solve the unique linear function from the environmental variables into the system state:

$$\bar{x} = A^{-1}B\,u, \tag{3.2}$$

so that one can define an explicit mapping matrix from u to $\bar{x}$

$$\phi^T = A^{-1}B. \tag{3.3}$$

However, the main motivation for the formulation in (3.1) is that one can formally extend the static model into a dynamic one. The formula (3.1) only characterizes the final visible balance in the system, but one has to remember that it is local operations only that exist — how can such uncoordinated local actions implement the global-level behaviors? Indeed, one needs to extend studies beyond the final balance, and take into account the dynamic behaviors caused by the imbalances.

Formula (3.1) can be interpreted as a balance of tensions determined by forces $A\bar{x}$ and Bu, caused by the system itself and by the environment, respectively. If the forces are not in balance, there is a drift. Assuming that the data structures are selected appropriately, so that $-A$ is stable (eigenvalues having negative real parts), one can define a dynamic model to characterize the tensions as

$$\frac{d\,x}{\gamma\,d\tau} = -Ax + Bu. \tag{3.4}$$

The parameter γ can be used for adjusting the time axis. The steady state equals that of (3.2), so that $\lim_{\tau\to\infty} x = \bar{x}$ for constant u. Because of linearity, this steady state is unique, no matter what was the initial state. Using the above construction, the static pattern has been transformed into a dynamic pattern — the observed equivalences are just emergent phenomena reflecting the underlying dynamic equilibrium.

How can such a genuine extension from a static model into a dynamic one be justified? It needs to be observed that there *must* exist such an inner structure beyond the surface. The seemingly static dependencies of the form (1.7) have to be basically dynamic equilibria systems so that the equality can be restored after disturbances: The actors, or the molecules in this case, do not know the "big picture", and it is the interactions among the molecules that provide for the tensions resulting in the tendency towards balance. It is assumed here that the mathematical model represents *what a system really does.* The model is not only mathematically appropriate, but it explains the actual mechanisms taking place in the chemical system that is getting towards balance after a transient.

What causes the dynamics, then? Thinking of the mindless actors in the system, the only reasonable explanation for the distributed behaviors is *diffusion.* It is the concentration gradients that only are visible at the local scale in a chemical system. So, interpreting (3.4) as a (negative) gradient, there has to exist an integral — a criterion that is being minimized. By integration with respect to the variable x, it is found that

$$\mathcal{J}(x,u) = \frac{1}{2}x^T Ax - x^T Bu \tag{3.5}$$

gives a mathematical "pattern" that characterizes the system in a yet another way. Note that by employing the dynamic systems understanding, it was possible to escape the limits of the "dead" formulation and turn an originally static problem into another, more interesting static form. Such an optimization-oriented view of systems as proposed above combines the two ways of seeing systems: The criterion itself represents the pattern view, whereas the optimization process represents the process view. Similarly, there is also connection to the philosophies: Whereas Heraclitus emphasized the processes, Plato tried to capture the "ideals", or the patterns beyond the changes.

Now one can conclude that the chemical balance system formally implements pattern matching of the form (2.12) as studied in the previous chapter, with variables being interpreted within the new structure:

$$\begin{aligned} J(x,u) &= \frac{1}{2}(u-\varphi x)^T W (u-\varphi x) \\ &= \frac{1}{2}x^T\varphi^T W\varphi x - x^T\varphi^T W u + \frac{1}{2}u^T W u, \end{aligned} \tag{3.6}$$

so that

$$J(x,u) = \mathcal{J}(x,u) + \frac{1}{2}u^T W u. \tag{3.7}$$

The two cost criteria (3.5) and (2.12) are equivalent what comes to the "tensions" imposed by them; constant factors (with respect to x) do not change the location of the minimum, nor the gradients, for given u. The correspondence between the cost criteria is reached when one defines the matrices as

$$\left\{ \begin{array}{lcl} A &=& \varphi^T W \varphi \\ B &=& \varphi^T W. \end{array} \right. \tag{3.8}$$

This connection between data structures is studied closer in Sec. 3.2.3. If (3.8) holds, one can see that all eigenvalues of A are non-negative, meaning that with such a selection the process (3.4) always remains stable.

Criterion (3.6) gives another view too see the same gradient-based minimization (3.4). When (3.6) is minimized using the steepest descent gradient approach, the continuous-time process implementing this minimization is

$$\frac{d\,x}{\gamma\, d\tau} = \varphi^T W (u - \varphi x). \tag{3.9}$$

It is the latter part $u - \varphi x$ that makes it possible to reach more sophisticated results in matching: For example, the adaptation can do the pattern matching even if the feature vectors in φ were non-orthogonal or unnormalized. This feedback structure will be studied later; now the key point is the basic structure of this formula (3.9). Whereas the matrix ϕ^T implements a mapping from the environmental variables u into the system variables $\bar{x}$, the feature matrix φ can be interpreted as an inverse mapping from the space of x into the space of u. Formally, simply for mathematical reasons, there must hold $\phi^T = \left(\varphi^T W \varphi\right)^{-1} \varphi^T W$, but more useful results can be found.

The effects φx and $\phi^T u$, or diffusion processes into and out from the system, eventually find their balance — it is not possible to determine the "original causes". One can even speak of a *holistic view* here. Because of pancausality, there exists a two-way connection: Changes in any variable causes changes in other variables, no matter whether the variable belongs to x or u. Just as the environmental variables can affect the system variables, the environment can be affected by the system. This two-way assumption blurs the traditional view of distinguishing between a system and its environment, there is no clear distinction between them. The "original" environment u is changed by x — but there does not exist any intact environment to start with. The vector u is the net effect of all accompanying subsystems, all of them together modifying their common substrate. A subsystem is an integral part of the whole, the

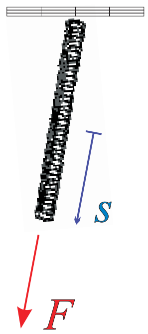


Figure 3.1: Prototypical spring being stretched

environment being a composition of subsystems. The whole system — indeed, the environment — consisting of a large number of subsystems can be dictated by none of the subsystems alone. The environment should not be seen as a distinct concept, or as something fundamentally intractable in the hierarchy of models. The deep connection between the mappings ϕ and φ is a key issue when trying to capture the behaviors of cybernetic systems.

However, such observations above have little value if the data structures ϕ, φ, and W (or A and B) cannot be determined. To attack this problem, a wider perspective is needed.

3.1.2 Elastic systems

Study the cost criterion (3.5) closer. It turns out that this cost criterion has a very familiar outlook, and employing new terminology, valuable intuitions are available. To see this, some facts need to be refreshed.

> Study a *spring* having the spring constant k (the spring can also be torsional, etc.). When the spring is stretched by an amount s because of an external force F (see Fig. 3.1), there are external and internal stored energies in the spring:
>
> - Due to the potential field: $W_{\text{ext}} = -\int_0^s F\,ds = -Fs$.
> - Due to the internal tensions: $W_{\text{int}} = \int_0^s ks\,ds = \frac{1}{2}ks^2$.
>
> This can be generalized, assuming that there are many forces, and many points being connected by springs, so that the internal tension between the points s_1 and s_2, for example, becomes
>
> $$W_{\text{int}}(s_1, s_2) = \frac{1}{2}k_{1,2}\left(s_1 - s_2\right)^2 = \frac{1}{2}k_{1,2}s_1^2 - k_{1,2}s_1s_2 + \frac{1}{2}k_{1,2}s_2^2.$$
>
> A matrix formulation is also possible for vectors s and F, when the interaction factors are collected in matrices A and B. It turns out that the expressions for potential energy components have familiar

outlooks:

$$W_{\text{int}}(s) = \frac{1}{2} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}^T A \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}, \tag{3.10}$$

and

$$W_{\text{ext}}(s, F) = - \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}^T B \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}. \tag{3.11}$$

For a moment, assume that vector u denotes *forces* acting in a (discretized) mechanical system, and x denotes the resulting *deformations.* Further, assume that A is interpreted as the *elasticity matrix* and B is *projection matrix* mapping the forces onto the deformation axes. Matrix A must be symmetric, and must be positive definite to represent stable structures sustaining external stresses — these conditions are fulfilled if (3.8) hold. Then, it turns out that (3.5) is the difference between the potential energies stored in the mechanical system. Principle of minimum potential (deformation) energy [19] states that a structure under pressure ends in minimum of this criterion, trying to exhaust the external force with minimum of internal deformations.

However, the same criterion can be seen to characterize all cybernetic balance systems. This means that in non-mechanical cybernetic systems, the same intuition concerning understanding of mechanical systems can be exploited. It does not matter what is the domain, and what is the physical interpretation of the "forces" u and of the "deformations" $\bar{x}$, the structure of the system behavior remains intact: As the system is "pressed", it yields in a more or less humble manner, but when the pressure is released, the original state is restored. Indeed, in chemical environments, this behavior is known as the *Le Chatelier principle:* If there is some disturbance acting on the system, the balance moves in such a direction where the effects are "eaten up". In this sense, one can generally speak of *elastic systems.*

In short: Neocybernetic systems are identical with elastic systems — systems that are characterized by dynamic equilibria rather than static equivalences. When rigid constraints are substituted by "soft" tensions, there is smoothness, and — by definition — local linearizability can be assumed also what comes to originally nonlinear models.

The effect of the environmental pressures on the system can be easily quantified: Just as in the case of a potential field, it is the product of the force and displacement that determines the change in potential energy. Similarly, regardless of the physical units of the variables, one can interpret the product $\bar{x}_i u_j$ in terms of *energy transferred from the environment into the system* through the pair of variables u_j and x_i. Correspondingly, if there are variables that can be interpreted as dissipative flows or rates, the energies are also effectively divided by time, so that it is some kind of *power* that is transferred. This concept deserves a name, or, actually, an old concept is renamed: In what follows, this "emergent level energy" is studied along the following definition:

> **Emergy** (a scalar dimensionless quantity) is the product of the (abstract) force and the corresponding (abstract) deformation.

As it turns out, this emergy is "information energy" that is the prerequisite for emergence of information structures. Emergy will here be a much more abstract thing and will have a broader scope than that used in [58].

Such energy flows have been studied before in more concrete terms in various contexts: *Bond Graphs* are used to model systems in terms of energies being transferred among system components [16]. It has been shown that this modeling strategy can be applied to a wide variety of tasks, so that this approach seems to be a rather general one. However, Bond Graphs are traditionally used for modeling different kinds of dissipative flows — and now the emphasis is on balances. Resulting models are very different.

However, it must be remembered that there is not only the effect from the external environment into the internal system — there is a symmetric two-way interaction that takes place. It is the matrices ϕ^T and φ that characterize the emergy transfer between the system and its environment. It is not only so that u should be seen as the "force" and $\bar{x}$ as the effect: $\bar{x}$ can be seen as the action and u as the reaction just as well. This duality makes it possible to tie the loose ends together.

3.1.3 Evolutionary fitness

It was mentioned above that the key challenge in this chapter is to determine the *goals of evolution.* Traditionally, one is facing paradoxes here: Remember that the layman intuition does not work. If the fitness criterion were the "maximum number of offspring", for example, there would only exist bacteria on earth. On the other hand, the "blind watchmaker" metaphor with random optimization [21] simply cannot be the mechanism beyond evolution.

Neocybernetic environment-orientedness suggests a criterion emphasizing some kind of *match with environment.* Indeed, applying the above discussion concerning energy/power transfer from the environment into the system and back, an intuitively appealing fitness criterion would be

> Maximize the average amount of emergy that is being transferred between the system and the environment.

No matter what is the physical manifestation of the environmental variables, a surviving system interprets them as resources, and exploits them as efficiently as possible. Note that it is not predetermined what should be done with the extracted energy: The metabolic products can change the environment to be further exploited. This makes it possible that evolutionary processes can proceed in many different ways — the relevance of the behaviors is later evaluated by the evolutionary selection. To begin with, the criterion is always the same — match with environment — no matter how some "master mind" would like the system to develop.

When there are resources available in the environment, it is also clever to utilize this abundance somehow. Typically, if the environmental "force" comes into a yeast cell in the form of glucose steps, for example, it is different kinds of metabolic products that can be produced: In some cases it can be the mannose-production path that outperforms others, producing new cells; in some other cases, heat production is to promote — meaning that reproduction and survival are competing goals. In each case, the assumption here is that the cell most efficiently exploiting the available energy prospers in the long run.

Following the above lines of thought, the momentary emergy traversing from the environmental variable j to the state variable i can be written as $\bar{x}_i u_j$, or, when written in a matrix form simultaneously for all variables, $\bar{x}u^T$. Similarly, the momentary emergy traversing from the state variable i to the environmental variable j can be written as $u_j \bar{x}_i$, or, when written simultaneously for all variables, $u\bar{x}^T$. If evolution proceeds in a consistent manner, the differences among such variable pairs should determine the growth rates of the corresponding links between the variables; when the mapping matrices ϕ^T and φ are defined as shown above, one can assume that a stochastic adaptation process takes place, the observations of prevailing variable levels determining the stochastic gradient direction:

$$\begin{cases} \dfrac{d\,\phi^T}{dt} & \propto \quad \bar{x}(t)u^T(t) \\ \dfrac{d\,\varphi}{dt} & \propto \quad u(t)\bar{x}^T(t). \end{cases} \tag{3.12}$$

However, note that the matrix elements cannot be explicitly localized in the system. When (structural) changes take place in the underlying system, it is constraints that are being added or modified, and these changes are reflected in the elements of ϕ^T and φ in more or less random ways. All changes in the underlying structure typically affect the mappings — but all of the changes affect them only little, at least if the number of components in the system is high. The high number of discrete parameters are projected onto the low-dimensional set of more or less smooth “emergent parameters”. When the discrete space of structures changes into a more continuous behavior of emergent parameters, more or less consistent evolutionary optimization becomes possible. What is more, the local optimizations are independent of each other — this makes it possible that various optimization processes can take place simultaneously, thus making the optimization a parallel process, relatively fast and robust. The time scales in (3.12) are much longer than in (3.4).

When looking at the formulas (3.3) and (3.12) together, for example, it is clear that such adaptation processes are unstable — high correlations between $\bar{x}_i$ and u_j eventually result in still higher correlations between them, thus making $\bar{x}_i$ (or u_j) grow without limit. Indeed, this adaptation principle is an extension of the *Hebbian learning rule,* where it is the correlation between the environmental signal in u_j and neuronal activity in $\bar{x}_i$ that has been shown to determine the synaptic adaptation in real neural cells [37].

There is a positive feedback in the adaptation law, and just as it is with the Hebbian neurons, the stability problem emerges if the trivial learning rule is applied (see [92]). Stabilization of the Hebbian learning model has been studied a lot — but, again, applying the neocybernetic simplicity ideal, one should not introduce new structures separately for stabilization purposes. For a moment, simply assume that $\bar{x}$ and u *for some reason* remain bounded; then it is reasonable to assume that the processes (3.12) find a fixed state, and the solution for this fixed state can be assumed to be such that the matrix elements ϕ_{ji} are relative to the correlations between $\bar{x}_i$ and u_j, or

$$\phi^T = q\,\mathrm{E}\left\{\bar{x}u^T\right\}, \tag{3.13}$$

and in the backward direction,

$$\varphi = b\,\mathrm{E}\left\{u\bar{x}^T\right\}. \tag{3.14}$$

If the dynamics of x is rather fast, so that the system can be assumed to always be in dynamic balance, one can substitute $\bar{x}$ with x in the above formulas (and also in the formulas that follow). Here, the parameters q and b are some constants; the role of these *coupling coefficients* is studied later. Similarly, the relevance of the observation (3.15), or the role of the system as a *mirror image* of the environment, will be discussed later. This means that the matrices ϕ and φ should become proportional to each other:

$$\varphi = \frac{b}{q}\,\phi. \tag{3.15}$$

As it turns out, these factors scale the signal levels in the system and in the environment. When interpreting (3.15), it is quite natural to think that *exploitation means exhaustion* — it is those elements u_j that contribute most in the determination of the values of $\bar{x}$ that become exhausted the most.

It needs to be recognized that the adaptation in the system according to (3.13) is completely local for any element in the matrices ϕ and φ even though the assumed goal of the evolutionary process is presented in a collective matrix format. It is interesting to note here that the expressions for ϕ and φ are essentially symmetric. Remember that it was Heraclitus who said "the way up and the way down are the same" — whatever he meant.

3.2 Towards self-organization

The key question in complex systems is that of self-organization: How can anything qualitatively new emerge from non-centralized operations. For a system to self-organize, it must first self-regulate. In this section, the issue of self-regulation is first studied, and the issue of self-organization after that.

The basic solution to regulation is negative feedback. However, now there are no explicit control structures available, and no organized communication or signal transfer infrastructure within the system: How to implement the feedback structures? Again, some background analysis is first in place.

3.2.1 Feedback through environment

The traditional approach to avoid explosions is to include non-idealities in the originally idealized models. For example, an originally linear system can become

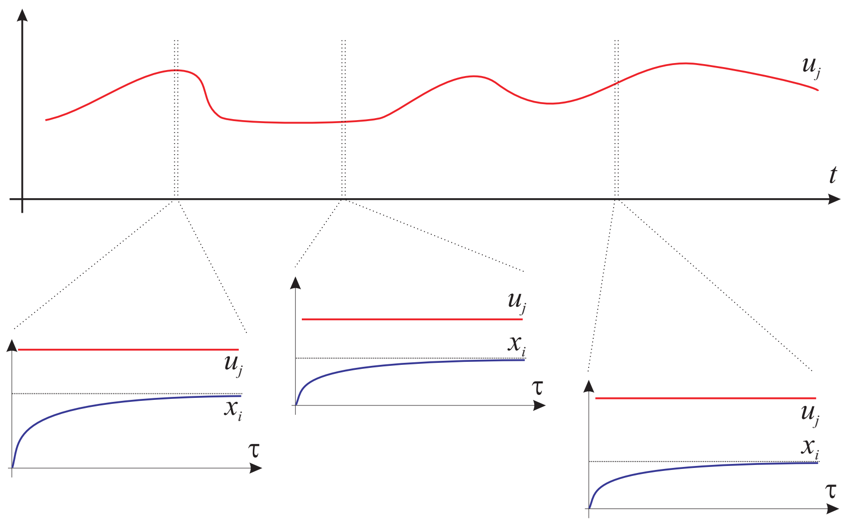


Figure 3.2: Illustration of two time scales. It is assumed that the dynamics of u (on the t scale) is much slower than that of x (τ scale)

stable if nonlinearities are added so that signals saturate. Here, non-idealities are again included in the model — however, these non-idealities are now located in an unorthodox place.

There are no unidirectional effects in real systems: Information flows cannot exist without physical flows that implement them. When emergy is being consumed by the system, this emergy is taken from the environment, or environmental "resources" are exhausted. To understand these mechanisms, study the pattern matching process (3.9). There are essentially two parts in this expression: First, in the front there is $\varphi^T W$ implementing parallel matching of data against the model, determining the directions of local diffusion processes; second, there is $u - \varphi x$ defining some kind of *virtual environment* that is being matched. The negative feedback structure $-\varphi x$ represents real material flow in from the system into the environment, the resources being exhausted. The changed environment becomes

$$\tilde{u} = \underbrace{u}_{\substack{\text{actual}\\ \text{environment}}} - \underbrace{\varphi x}_{\text{feedback}} . \tag{3.16}$$

The system never sees the original u but only the distorted $\tilde{u}$, where the momentary emergy consumption in the system, or φx, is taken into account. Clearly, as the environment affects the system and the system affects the environment, there exists a feedback structure; again, one is interested in the final balance after transients:

$$\bar{u} = u - \varphi \bar{x}. \tag{3.17}$$

Later on, real-life realism will be applied: Only $\bar{u}$ is visible, never u itself. The matrix ϕ^T is redefined here: It stands for the mapping from the effective environment to the state, however this environment is manifested — in this feedback case meaning that $\bar{x} = \phi^T \bar{u}$.

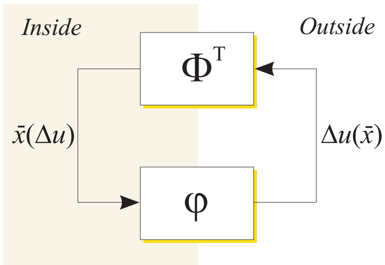


Figure 3.3: The algebraic loop between the environment and the system

Because the environment is disturbed by the system, the setting is nonideal, but this nonideality makes new functionalities possible — like self-organization, as shown in the next section. But the key issue here is that this negative feedback keeps the system in balance and signals bounded, as was assumed in the previous section. The feedback structure is implicit, through the environment, and the effects of this feedback will be studied below. To start with, no assumptions like (3.14) are made — φ is an arbitrary $m \times n$ mapping matrix.

When studying the steady state, there is efficiently an algebraic loop in the system (see Fig. 3.3), and this means that this structure has peculiar properties. Multiplying (3.17) from the right by $\bar{x}^T$, taking expectations, and reordering the terms, one receives

$$\mathrm{E}\{(u - \bar{u})\bar{x}^T\}\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} = \varphi, \tag{3.18}$$

so that, when one defines a quantity for measuring the discrepancy between the undisturbed open-loop environment and the disturbed closed-loop environment,

$$\Delta u = u - \bar{u}, \tag{3.19}$$

the expression (3.17) can be written in the form

$$\Delta u = \mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\bar{x}. \tag{3.20}$$

Variables in $\bar{x}$ and Δu are mutually connected, they vary hand in hand, but together representing the same mapping as φ, but in terms of observation data, helping to see another view of the system properties. Indeed, this Δu can be seen as a "loop invariant" that helps to see properties of the feedback loop, and it turns out to offer a way to reach simplified analysis of the signals. Because Δu assumedly linearly dependent of u, one can interpret this variable as the actual input driving the whole loop, so that there exists a mapping Φ^T

$$\bar{x} = \Phi^T\,\Delta u. \tag{3.21}$$

Assuming that the feedback can implement stabilization, the system in Fig. 3.3 will search a balance so that

$$\bar{x} = \Phi^T\varphi\,\bar{x}. \tag{3.22}$$

To have not only trivial solutions (meaning $\bar{x} \equiv 0$), there must hold

$$\Phi^T\varphi = I_n, \tag{3.23}$$

so that the feedforward and feedback mappings have to be mutually orthogonal. This is a very stringent constraint, and it essentially determines the properties of the feedforward matrix Φ. Here, to determine Φ, assume symmetry with (3.18), and make the following attempt and study where it leads to:

$$\Phi^T = \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}\Delta u^T\}. \tag{3.24}$$

3.2.2 Back to principal subspace

Above, the balances of x were studied as the environment u was assumed fixed. However, to reach interesting results, the neocybernetic principles need to be exploited again: It is assumed that there exist various levels of seeing the system, and at each of the levels, the balances are exploited. Specially, see Fig. 3.2: Whereas u was assumed to remain constant this far, it only has much slower dynamics than x, and on the wider scale, the environment changes. But assuming stationarity of the environment, or balance on the higher scale, so that u has fixed statistical properties, one can find a "balance model of balances". A truly cybernetic model is a *second-order balance model,* or a *higher-order balance model* over the variations in the system — at these levels beyond the trivial first level balance, one can reach stronger views to see the systems, including *self-organization,* as shown below.

So, assume that dynamics of u is essentially slower than that of x and study the statistical properties over the range of $\bar{x}$, and, specially, construct the covariance matrix of it. From (3.24) one has

$$\bar{x}\bar{x}^T = \mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}\mathrm{E}\left\{\bar{x}\Delta u^T\right\}\ \Delta u\Delta u^T\ \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}. \tag{3.25}$$

When applying expectation operator on both sides,

$$\mathrm{E}\left\{\bar{x}\bar{x}^T\right\} = \mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}\mathrm{E}\left\{\bar{x}\Delta u^T\right\}\ \mathrm{E}\left\{\Delta u\Delta u^T\right\}\ \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}.$$

Multiply these from left and from right by $\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}$:

$$\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^3 = \mathrm{E}\left\{\bar{x}\Delta u^T\right\}\ \mathrm{E}\left\{\Delta u\Delta u^T\right\}\ \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T, \tag{3.26}$$

and, when observing the nature of Φ, this can be written

$$\left(\Phi^T\mathrm{E}\left\{\Delta u\Delta u^T\right\}\Phi\right)^3 = \Phi^T\mathrm{E}\left\{\Delta u\Delta u^T\right\}^3\Phi. \tag{3.27}$$

If $n = m$, any orthogonal matrix $\Phi^T = \Phi^{-1}$ will do; however, if $n < m$, so that x is lower-dimensional than u, the solution to the above expression is non-trivial (see [92]: Report 144, "Hebbian Neuron Grids: System Theoretic Approach", pages 12–15). It turns our that *any subset of the principal component axes of the data* Δu *can be selected to constitute* Φ*,* that is, the columns Φ_i can be any n of the m covariance matrix eigenvectors θ_j of this data. Further, these basis vectors can be mixed, so that $\Phi = \theta D$, where D is any orthogonal $n \times n$ matrix[1], so that $D^T = D^{-1}$. In any case, there holds

$$\Phi^T\Phi = I_n. \tag{3.28}$$

[1]Note that there is an *error* in that report in [92]: The matrix D is not whatever invertible matrix, it must be orthogonal (as becomes evident when going through the proof therein)

Now, return to the assumption in (3.24) — indeed, the above selection for Φ seems to fulfill the orthogonality claim (3.23):

$$
\begin{array}{rcl}
\Phi^T\varphi & = & \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}\Delta u^T\}\,\mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
& = & \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\Phi^T\mathrm{E}\{\Delta u\Delta u^T\}\,\mathrm{E}\{\Delta u\Delta u^T\}^T\Phi\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
& = & \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\Phi^T\mathrm{E}\{\Delta u\Delta u^T\}^2\,\Phi\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
& = & \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\mathrm{E}\{\bar{x}\bar{x}^T\}^2\,\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
& = & I_n.
\end{array}
\tag{3.29}
$$

The above derivations show that any set of covariance matrix eigenvectors can be selected in Φ. However, in practice it is not whatever combination of vectors θ_j that can be selected: Some solutions are *unstable* when applying the iterative adaptation strategies. Indeed, following the lines of thought shown in [92], the only stable and thus relevant solution is such where it is the n most significant eigenvectors (as revealed by the corresponding eigenvalues) that constitute the matrix Φ in convergent systems. This means that the system implements *principal subspace analysis* for input data. Because of the mixing matrix D, the result is not unique in the sense of principal components, but the subspace spanned by them is identical, and exactly the same amount of input data variation is captured. Specially, if there were some further exploitation of the latent variables $\bar{x}$, reconstructions of $\hat{y}$ would be equally accurate no matter whether the principal components or the principal subspace only were used.

The above derivations apply to all feedback matrices φ: The system signals adapt to fulfill the equation (3.18). The results only apply to the subspace spanned by φ — that is, in the subspace where there is variation in Δu caused by the feedback — and within that subspace, the structure of maximum variation is found. If φ is adaptive and selected applying the evolutionary strategy, so that $\varphi^T = b\mathrm{E}\{\bar{x}\bar{u}^T\}$, it is the principal subspace of u that is spanned. These issues will be studied later.

Now one can conclude that completely local operations result in non-trivial structures that are meaningful on the global scale: Competitive learning without any structural constraints results in self-regulation (balance) and self-organization (in terms of principal subspace). Feedback through the environment, or competition for the resources, results in stabilization and organization of the system.

3.2.3 Closer look at the cost criteria

When comparing to (3.3) to (3.24), and when u in the formulas is substituted with Δu, one can see that an appropriate connection between the data structures is reached when one selects the matrices so that

$$
\left\{
\begin{array}{rcl}
A & = & \mathrm{E}\left\{\bar{x}\bar{x}^T\right\} \\
B & = & \mathrm{E}\left\{\bar{x}\Delta u^T\right\}.
\end{array}
\right.
\tag{3.30}
$$

As presented in [92], essentially the same formulas were found in the neuronal system applying not only "Hebbian learning", but together with the "anti-Hebbian" structures, where the feedbacks were explicitly implemented. When the feedback through the environment is taken into account, simpler structures suffice, and the results are the same. However, there is a difference: Whereas the explicitly implemented feedback structures analyze the original undisturbed environment, the feedbacks implemented through the environment analyze the disturbances in the environment. These differences between open-loop environment and closed-loop environment are measurable only after adaptation, Δu substituting the original u in analyses. The model with explicit feedback is not completely based on local information: There the matrix ϕ implements a mapping from u onto $\bar{x}$, essentially assuming that the feedback is implemented without affecting the environment itself. Such a feedback scheme is possible in systems where the actors are "intelligent agents" that are capable of seeing the environment in a wider perspective, as studied in the next chapter.

Yet another conclusion is in place here: Comparing expressions (3.8) and (3.30), it turns out that to avoid contradictions, one has to choose $W = \mathrm{E}\{\Delta u \Delta u^T\}$. If the feedback is explicit, on the other hand, the weighting matrix is $W = \mathrm{E}\{uu^T\}$. The implicit data weighting is also identical with that proposed in the context of emergent models. The technical manipulations in the previous chapter are essentially an appropriate way to characterize the behaviors also in the locally controlled, real (but idealized) system:

> Neocybernetic system implements the emergent model structure. The locally controlled system carries out modeling of the environment u applying principal subspace based feature extraction (slow process of determining ϕ) and pattern matching (fast process of determining $\bar{x}$).

Having compact formulations for the matrices, the cost criteria can also be studied closer. Defining $\mathcal{J}(u) = \mathcal{J}(\bar{x}, u)$, from (3.5) one has, assuming that there holds (3.24,

$$\begin{aligned}
\mathcal{J}(u) &= \frac{1}{2}\,\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x} - \bar{x}^T\mathrm{E}\{\bar{x}\Delta u^T\}\Delta u \\
&= \frac{1}{2}\,\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x} - \Delta u^T\mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}\Delta u^T\}\Delta u \\
&= \frac{1}{2}\,\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x} \\
&\quad - \underbrace{\Delta u^T\mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}}_{\bar{x}^T}\mathrm{E}\{\bar{x}\bar{x}^T\}\underbrace{\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}u^T\}\Delta u}_{\bar{x}} \\
&= -\frac{1}{2}\,\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x},
\end{aligned}$$

so that the *average* of the criterion can be written as

$$\begin{aligned}
\mathrm{E}\{\mathrm{trace}\{\mathcal{J}(u)\}\} &= -\frac{1}{2}\,\mathrm{E}\left\{\mathrm{trace}\left\{\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x}\right\}\right\} \\
&= -\frac{1}{2}\,\mathrm{E}\left\{\mathrm{trace}\left\{\bar{x}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\right\}\right\} \\
&= -\frac{1}{2}\,\mathrm{trace}\left\{\mathrm{E}\left\{\bar{x}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\right\}\right\} \\
&= -\frac{1}{2}\,\mathrm{trace}\left\{\mathrm{E}\{\bar{x}\bar{x}^T\}^2\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{n}\lambda_i^2.
\end{aligned} \tag{3.31}$$

The above simplification comes from the linearity of the operators, $\mathrm{trace}\,\{\mathrm{E}\{\cdot\}\} = \mathrm{E}\,\{\mathrm{trace}\{\cdot\}\}$, and from the properties of matrix trace: Trace it is the sum of the diagonal elements, and simultaneously it is the sum of the matrix eigenvalues; for scalars there is naturally no effect. What is more, matrices within trace can be rotated, that is, $\mathrm{trace}\{M_1 M_2\} = \mathrm{trace}\{M_2 M_1\}$, if the matrices M_1 and M_2 are appropriately compatible. The above result means that the completely adapted system maximizes the sum of the n most significant eigenvalue squares as seen from within the system. Using the other criterion, the optimum reaches $\mathrm{E}\{J(u)\} = \sum_{j=n+1}^{m} \lambda_j^2$.

It has to be kept in mind that if the feedbacks are implemented through the environment, the eigenvalues λ_i are eigenvalues of $\mathrm{E}\{\Delta u \Delta u^T\}$. They are eigenvalues of $\mathrm{E}\{uu^T\}$ only if the feedbacks are implemented actively by some intelligent agent (as studied in later chapters).

3.2.4 Making it local

The above theoretical derivations were interesting, giving qualitative understanding of the properties of the feedback loop, but they were applicable only for the global scale analyses: From the point of view of the system, Δu is not known, as the original undisturbed u cannot be seen without disturbing it. So, from now on, assume that the system only sees the real, virtual environment as disturbed by the feedbacks, and, according to (3.13), define[2]

$$\bar{x} = \phi^T \bar{u}, \tag{3.35}$$

where $\phi^T = q\mathrm{E}\{\bar{x}\bar{u}^T\}$. Now it is the really measurable environment, as manifested in $\bar{u}$, that is only involved in local calculations. As it is the feedback that supplies for the basic functionality of a cybernetic system, spanning the principal subspace of the data, it is the role of the learning to make this data represent the external environment u as manifested in $\bar{u}$. There are two main functionalities in the studied system structure: Feedback implements principal subspace analysis, and adaptation in the form (3.13) and (3.14) implements match with environment, so that it is the signals Δu, and simultaneously the original u, that determine this principal subspace. Going towards maximum variation spans the principal subspace in the data when the latent variables are kept linearly independent.

[2]How is (3.35) related to (3.24), how can they represent the same system — specially, where does the inverse covariance matrix $\mathrm{E}\{\bar{x}\bar{x}^T\}$ emerge in the formula? To have intuition on this, note that

$$\begin{array}{rcl} \phi^T & = & q\,\mathrm{E}\{\bar{x}\bar{u}^T\} \\ & = & q\,\mathrm{E}\{\bar{x}(u - b/q\,\phi\bar{x})^T\} \\ & = & q\,\mathrm{E}\{\bar{x}u^T\} - b\,\mathrm{E}\{\bar{x}\bar{x}^T\}\,\phi^T, \end{array} \tag{3.32}$$

and when solving this,

$$\phi^T = \left(\mathrm{E}\{\bar{x}\bar{x}^T\} + \frac{1}{b}\,I_n\right)^{-1} \frac{q}{b}\mathrm{E}\{\bar{x}u^T\}. \tag{3.33}$$

When letting b grow, the required functional structure emerges:

$$\phi^T = \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\frac{q}{b}\mathrm{E}\{\bar{x}u^T\}. \tag{3.34}$$

The signals $\bar{x}$ and $\bar{u}$, as defined as in (3.35), have peculiar properties. For example, multiplying (3.35) from the right by $\bar{x}^T$ and taking expectation, one has an expression for the latent vector covariance:

$$\mathrm{E}\{\bar{x}\bar{x}^T\} = q\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T. \tag{3.36}$$

This holds *if* the latent variables x_i do not fade away altogether (or explode). These issues are studied later — however, here it is assumed that the system is strictly cybernetic, all latent variables are occupied, and, for example, the matrix $\mathrm{E}\{\bar{x}\bar{x}^T\}$ remains invertible. On the other hand, multiplying (3.35) from the right by $\bar{u}^T$ and taking expectation, one has

$$\mathrm{E}\{\bar{x}\bar{u}^T\} = q\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{u}\bar{u}^T\}. \tag{3.37}$$

Substituting this in (3.36),

$$\mathrm{E}\{\bar{x}\bar{x}^T\} = q^2\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{u}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T, \tag{3.38}$$

or

$$\begin{aligned} \frac{1}{q}\,I_n &= \sqrt{q}\,\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\mathrm{E}\{\bar{x}\bar{u}^T\}\,\mathrm{E}\{\bar{u}\bar{u}^T\}\,\mathrm{E}\{\bar{x}\bar{u}^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\sqrt{q} \\ &= \bar{\theta}'^T\,\mathrm{E}\{\bar{u}\bar{u}^T\}\,\bar{\theta}', \end{aligned}$$

where

$$\bar{\theta}'^T = \sqrt{q}\,\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\mathrm{E}\{\bar{x}\bar{u}^T\}. \tag{3.39}$$

From (3.36), it is evident that there holds[3]

$$\bar{\theta}'^T\bar{\theta}' = I_n. \tag{3.40}$$

This means that the columns in $\bar{\theta}'$ span the subspace determined by n of the principal components of $\mathrm{E}\{\bar{u}\bar{u}^T\}$, so that $th\bar{e}ta' = \bar{\theta}D$, where $th\bar{e}ta$ is a matrix containing n of the covariance matrix eigenvectors, and D is some orthogonal matrix; as in Section 3.2.2, it can be assumed that this is the principal subspace spanned by the n most significant of them (this claim is confirmed by simulations). All eigenvalues $\bar{\lambda}_j$ in the *closed loop* equal $1/q$.

Assume that the coupling coefficients q_i vary between latent variables, so that one has $\phi^T = Q\mathrm{E}\{\bar{x}\bar{u}^T\}$ for some diagonal coupling matrix Q. Following the above guidelines, it is easy to see that the matrix of eigenvalues for $\mathrm{E}\{\bar{u}\bar{u}^T\}$ becomes Q^{-1}. What is more interesting, is that one can derive for the symmetric matrix $\mathrm{E}\{\bar{x}\bar{x}^T\}$ two expressions: Simultaneously there holds $\mathrm{E}\{\bar{x}\bar{x}^T\} = Q\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T$ and $\mathrm{E}\{\bar{x}\bar{x}^T\} = \mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^TQ$. For non-trivial Q, and if the eigenvalues are distinct, this can only hold if latent vector covariance is

[3] The property (3.36) has also practical consequences. Recognizing that the Hessian $d^2J(x)/dxdx^T$ of the criterion (3.5) becomes a scaled identity matrix, it is evident that the originally first-order convergence properties of the gradient descent process (3.4) change into second-order dynamics, the process becoming an implementation of Newton method towards reaching the balance $\bar{x}$ after a transient

diagonal; what is more, the vectors in $\bar{\theta}^T = \sqrt{Q}\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\mathrm{E}\{\bar{x}\bar{u}^T\}$ now not only span the principal subspace, but they are the PCA basis vectors themselves (basis vectors not necessarily ordered in the order of significance). This means that the modes become separated from each other if they are coupled to the environment in different degrees.

The eigenvectors of u are the same as those of $\bar{u}$, but the eigenvalues are evidently not. Now study how the realizable mapping ϕ^T affects on the virtual mapping between u and $\bar{x}$. From (3.35) one has

$$\bar{x} = \sqrt{q}\,\mathrm{E}\{\bar{x}\bar{u}^T\}\,(u - b\varphi\bar{x})\,, \tag{3.41}$$

and, when solving for $\bar{x}$ and when recognizing (3.36),

$$\begin{array}{rcl} \bar{x} & = & \left(I_n + bq\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T\right)^{-1} q\mathrm{E}\{\bar{x}\bar{u}^T\}\,u \\ & = & \left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right)^{-1} q\mathrm{E}\{\bar{x}\bar{u}^T\}\,u, \end{array} \tag{3.42}$$

so that

$$\left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right)\,\mathrm{E}\{\bar{x}\bar{x}^T\}\,\left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right) = q^2\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{uu^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T,$$

or, utilizing (3.39),

$$\begin{array}{l} \left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right)^2 \\ \qquad = q\,\sqrt{q}\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\,\mathrm{E}\{\bar{x}\bar{u}^T\}\,\mathrm{E}\{uu^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\sqrt{q} \\ \qquad = q\,\bar{\theta}^T\,\mathrm{E}\{uu^T\}\,\bar{\theta}. \end{array}$$

This comes from the fact that $M\,f(M) = f(M)\,M$ for a square matrix M and a function f that is defined in terms of a matrix power series. From the fact that the eigenvectors $\bar{\theta}_j$ of $\mathrm{E}\{\bar{u}\bar{u}^T\}$ are also eigenvectors of $\mathrm{E}\{uu^T\}$, one has

$$\mathrm{E}\{\bar{x}\bar{x}^T\} = \frac{\sqrt{q}}{b}\,\bar{\theta}^T\mathrm{E}\{uu^T\}^{1/2}\bar{\theta} - \frac{1}{b}\,I_n. \tag{3.43}$$

The eigenvalues of $\mathrm{E}\{\bar{x}\bar{x}^T\}$ also can be expressed in terms of the n most significant eigenvalues λ_j of the undisturbed $\mathrm{E}\{uu^T\}$:

$$\frac{\sqrt{q\lambda_j} - 1}{b}. \tag{3.44}$$

As compared to the discussion in Section 3.2.2, the refined model has qualitatively very different properties: Whereas in the nominal principal component model the variation in input is maximally inherited by the latent structure, so that $\sum_{i=1}^n \mathrm{E}\{\bar{x}_{ii}^2\} = \sum_{j=1}^n \lambda_j$, now there is loss of variation within the system.

3.3 Analysis of elasticity

This section concludes the mathematical analysis of the generic neocybernetic framework. Intuitively, it is elasticity that will pop up every now and then in the subsequent analyses, and the conceptually farthest-ranging consequences come from the rigidity of the feedback structure: The environment changes its outlook because of the systems in it.

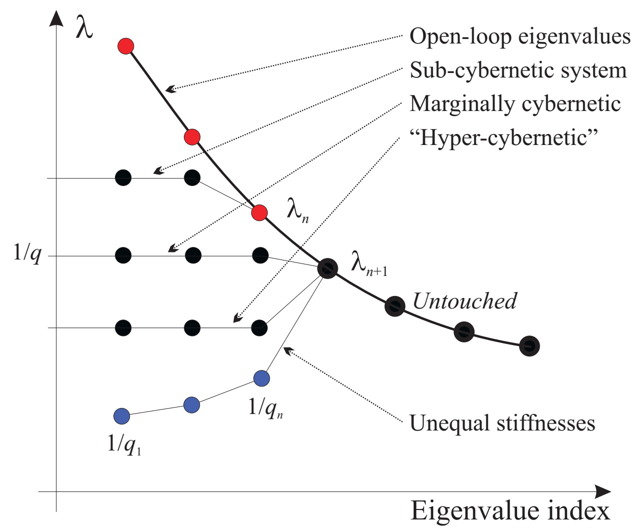


Figure 3.4: Schematic view of the cybernetic eigenvalues

3.3.1 Balance between system and environment

Because of the cybernetic scaling in the form $\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}$, the latent variables cannot go to zero, and a balance is found where the opposing drifting effects are compensated. In the directions dictated by the mapping matrices ϕ and φ (or $\mathrm{E}\{\bar{x}\bar{u}^T\}^T$), there is loss of excitation in the environment, as studied in Section 3.2.4, so that equalization of environmental variation takes place. This kind of "trivialization" of the environment is implemented not only through adaptation in the system, but also through changes in the environment. These results concerning "constant elasticity" are of extreme importance and they will be studied later.

It is also so that the environmental variation is suppressed, but simultaneously it is inherited by the system manipulating the environment. To reach such cybernetic situation where all n latent variables remain occupied, from (3.44) it is evident that there must hold $q\lambda_n > 1$. This means that there has to exist enough excitation to invoke the system, and make the adaptation process *without* the feedback unstable.

It is also possible to have separate values for q_i and b_i in different feedback loops, represented by different latent variables x_i, so that mappings $\phi^T = Q\mathrm{E}\{\bar{x}\bar{u}^T\}$ and $\varphi^T = B\mathrm{E}\{\bar{x}\bar{u}^T\}$ become "species-specific":

$$Q = \begin{pmatrix} q_1 & & 0 \\ & \ddots & \\ 0 & & q_n \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_n \end{pmatrix}. \tag{3.45}$$

Then it is not the principal subspace only that is constructed in the cybernetic process — it turns out that different eigenvalues are localized, and $\mathrm{E}\{\bar{x}\bar{x}^T\}$ becomes diagonal. The remaining covariance matrix corresponding to the cybernetic modes in the environment is, as projected onto the n-dimensional principal subspace,

$$\begin{pmatrix} \frac{1}{q_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{q_n} \end{pmatrix}, \tag{3.46}$$

and the induced covariance matrix of the cybernetic modes in the system is

$$\begin{pmatrix} \frac{\sqrt{q_1 \lambda_{j(1)}}-1}{b_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\sqrt{q_n \lambda_{j(n)}}-1}{b_n} \end{pmatrix}. \tag{3.47}$$

Here, notation $j(i)$ means that any permutation of the n most significant eigenvalues of $\mathrm{E}\{uu^T\}$ is possible. It turns out that all cross-correlations among system variables are eliminated, $\mathrm{E}\{\bar{x}\bar{x}^T\}$ being diagonal; the covariance $\mathrm{E}\{\bar{u}\bar{u}^T\}$ is not diagonal, though. It also turns out that when the feedback is implemented through the environment, one can have $n = m$ without losing the cybernetic properties of the system. To be sure that all modes are cybernetic, there must hold

$$q_i \lambda_n > 1. \tag{3.48}$$

In Figure 3.4, such situation where all modes fulfill the above constraint, is called (marginally) cybernetic, whereas cases where the "coupling" is too weak is called "sub-cybernetic". At least for some of the latent variables, in the closed-loop system one has $0 = q_i \mathrm{E}\{\bar{x}_i \bar{u}^T\}\bar{u}$ — the simplest solution for this is the trivial $\bar{x}_i \equiv 0$ and $\mathrm{E}\{\bar{x}_i \bar{u}^T\} = 0$, and the latent variable can fade away altogether. In real, converged systems, it can also be assumed that existent, non-vanishing latent structures cannot be sub-cybernetic. Further, looking at Fig. 3.4: If the (visible) variation structure changes so that the ordering of the eigenvalues becomes blurred, less significant variation directions outweighing the originally more significant ones, the situation is called "hyper-cybernetic". Note that the system still sees the original variation in u rather than the compensated in $\bar{u}$, so that there are no convergence problems however high the values of q_i are.

The parameters q_i and b_i remain free design parameters: Different kinds of system / environment combinations are instantiated for different selections of them, all of them equally valid, as long as (3.48) is fulfilled. Now it is possible to interpret these coupling coefficients in intuitive terms:

- **Stiffness ratio** q_i determines how tightly connected the system is into its environment, and how aggressively the system affects the environment, directly determining how "rigid" the corresponding direction in the data space is.

- **Dissipation rate** b_i determines how efficiently variation on the lower level (environment) is transferred onto the higher level (system itself). The non-transferred portion can be seen as loss of resources — see next chapters for closer analyses.

To assure cybernetic operation of the system, one can also make q_i adaptive. For example, local manipulations only are needed if one selects ($\nu > 1$ being some constant)

$$q_i = \frac{\nu}{\mathrm{E}\{\bar{x}_i^2\}}. \tag{3.49}$$

However, for a strictly cybernetic variable the above expression is automatically fulfilled, as the variance of the variable is relative to the inverse of the coupling factor, and other kinds of adaptation strategies for q_i can be proposed (see chapter 6).

3.3.2 Power of analogies

When applying linear models, the number of available structures is rather limited – indeed, there exist more systems than there are models. This idea has been applied routinely: Complicated systems are visualized in terms of more familiar systems with the same dynamics. In the presence of modern simulation tools, this kind of lumped parameter simplifications seem somewhat outdated — however, in the case of really complicated distributed parameter systems, such analogies may have reincarnation.

Mechanical associations

The original intuition concerning mechanical deformable systems in Sec. 3.1.2 can be extended. Think of a steel plate: If there are external forces acting on the plate, there is a continuum of smooth deformations on the surface. The plate is a distributed parameter system, but the distinct actors are like "probes", discretizing the state space, channeling the infinite-dimensional system onto the finite set of variables. Not all forces affecting the system can be detected, and not all deformations can be compensated — but what comes to the *visible* phenomena, projected through the observables onto the realm of concrete numbers, they can be mastered in the neocybernetic framework, exploiting the above observations: The variation structures become restructured.

The infinite complexity of the environment (the "forces") are mapped onto the measurements ("deformations"). A special case — but typical in practical systems — is the distributed case where individual observations and feedbacks are paired: Only the local environment can be observed, and it is this local environment that is mainly affected by the corresponding feedback. Now the structure of the environment is determined by this setup: No matter where the sensor/actuator pairs are located, the deformations in those locations become equalized and separated. The system variables are *a priori* fixed, and the whole infinite-dimensional "world" becomes anchored by the sensor/actuators.

In a mechanical system, such sensor-actuators are naturally separated in space. However, in more abstract systems, separation is not about spatial but about more complicated (higher-dimensional) dependency structures. The mechanical analogy makes the high-dimensional domain fields better graspable, projecting the wealth of simultaneous variables into the wealth of locations along the hypothetical plate, the interpretations of the semantics-loaded variables being made

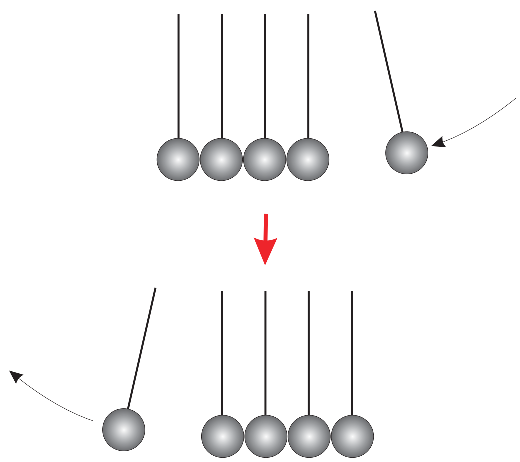


Figure 3.5: Maximum energy transfer is reached when impedances match. The physical units of "impedance" can vary (here it is the *masses* that need to be matched)

commeasurable — everything is only about "forces" and "deformations". It can be said that *dimensional complexity changes into spatial diversity.* These issues are studied closer in the subsequent chapters.

Electrical understanding

Another type of analogues are also routinely constructed: One can select *electrical current* and *voltage* rather than force and deformation. The external forces change to electrical loads disturbing the system: The deformation is the voltage drop, and the compensating action is the increased current supply (or vice versa). Traditionally, the non-idealities (output voltage drops when current is used) make it difficult to study interconnected groups of systems — the information flow is not unidirectional — but now the neocybernetic framework makes it possible to *exploit* these underlying feedbacks, even though they are implicit. Applying the distributed parameter framework instead the traditional lumped parameter one, one can reach again new intuitions, getting rid of SISO thinking.

The electric analogy makes it possible to extend the inner-system discussions onto the environmental level, to inter-system studies. When there are many connected systems interacting, one subsystem exhausting emergy supplied by the other subsystems — or providing emergy for the others, or transferring emergy between them — the criterion for system fitness can be based on the power transmission capabilities among the systems. And it is the product of current and voltage that has the unit of power, so that exactly the above discussions apply. Only the intuitions change: Now one can utilize the *inter-system* understanding supplied by electrical systems. Remember that the maximum throughput without "ringing" between electrical systems is reached when there is *impedance matching:* The output impedance in the former system and the input impedance of the latter one should be equal, otherwise not all of the power goes through but bounces back (however, in a non-mechanical/non-electrical system, there is not necessarily inertia, and no oscillatory tendency). This same bouncing metaphor can be applied efficiently also in non-electrical environments — the variables can have different interpretations but the qualitative behaviors remain the same (for example, see Fig. 3.5). It is not only local agent-level optimization that results in global system-level optimization, it is local system-level optimization that finally results in global environment-level optimization. These intuitions will be exploited in the following chapter.

Again, it is natural to study systems where the pairs of input and output variables are localized. Assume that $m = n$ and the variables are coupled as pairs, that is, the mapping matrix ϕ is diagonal, and $\bar{u}_i$ and $\bar{x}_i$ go together. The electrical analogy makes it possible to interpret the role of the coupling coefficients q_i in the formulas in a new way. As it is this parameter that connects the input (voltage) and the output (current) for cybernetic systems, $\bar{x}_i = q_i \mathrm{E}\{\bar{x}_i \bar{u}_i\}\, \bar{u}_i$, it is $Z_i = 1/q_i \mathrm{E}\{\bar{x}_i \bar{u}_i\}$ that explicitly stands for impedance. It is also so that in an evolutionary surviving environment the corresponding impedances have to become equal. This means that there is yet another iterative optimization loop — this time not within one system, but between all pairs of systems within an environment.

The field of electrical engineering is a highly sophisticated branch of powerful mathematics, and there developed conceptual tools can directly be exploited also in the analysis of cybernetic systems. There are some extensions that are needed:

- This far, only real-valued variables have been seen reasonable, and the models have been constructed accordingly. However, if transpositions are always changed to Hermitean matrices, so that in addition to transposing the matrices the elements are also complex conjugated, all of the above analyses can directly be extended to complex domain, so that all variables and matrices can consist of real and imaginary parts.
- What is more, only scalar variables have this far been reasonable. However, the variables can be extended to function domain: The variables can be parameterized, so that the constructed models and data structures remain functions of these parameters. So, if the extra parameter is the angular frequency ω, the analyses can be carried out in frequency domain — and then one needs the complex variables.

The above extensions make it possible to study dynamic phenomena by applying essentially the same formulas. Impedances $Z_i(s)$ can be interpreted in terms of *dynamic filters* between Laplace-transformed signals $\bar{U}_i(s)$ and $\bar{X}_i(s)$, being transfer functions of the complex variable s. The explicit spectra of $Z_i(s)$ can be found for values $s = \mathrm{i}\omega$, and the inverse transforms as $\bar{x}(t) = \mathcal{L}^{-1}\bar{X}(s)$. This means that it is not only the final balance that can be studied in the neocybernetic framework but also the stationary non-balance phenomena — and, indeed, dynamic models are most appropriate for real life systems, where understanding of how they behave during transients is very relevant. To reach best possible power transfer it is also these frequency-domain functions $Z_i(s)$ that need to match in neighboring systems. If the system can efficiently affect its environment, there is an iteration process where all systems constituting the environment adapt to find the common balance, that is, the objective $Z_i(s)$ are not given *a priori*.

Further, if the whole environment is evolutionarily optimal, it is the above observations that characterize the behaviors: The matrix $\mathrm{E}\{\bar{X}(s)\bar{X}^H(s)\}$ can be interpreted as a matrix of *autospectra* and *cross-spectra* for signals in $\bar{X}(s)$. Again, there are surprises:

> Comparing to (3.36), it is evident that the functions $q_i(s)$ must be selected so that they can be interpreted as autospectra, so that $q_i(s) = q_i'(s)q_i'(-s)$ for some valid transfer function $q_i'(s)$. The power spectrum must be real (and non-negative) for all frequencies, meaning that there exist no phase properties present in such spectra. This means that in $q_i(\mathrm{i}\omega)$ there are no phase properties, the transfer function containing no actual dynamics — meaning that $q_i(\mathrm{i}\omega)$ must have the same value for all frequencies: Coupling coefficient q_i must be constant and real.

The only remaining degrees of freedom in this extremum is the values of the interaction constants q_i and b_i. And, indeed, such questions are very relevant in everyday systems — or, actually, they may be relevant actually to "everything" ... see chapter 9 for more discussion. The system-internal frequency-domain considerations are elaborated on from another perspective again in chapter 5.

3.3.3 Applications in engineering systems

What happens if the evolutionary adaptation scheme is applied in technical systems? The discussions above were idealized, assuming absolute evolutionary optimality (as defined in terms of emergy transfer). However, to exploit the intuitions in real-life systems where the assumptions about maximum coupling with the environment do not hold, some more analysis is needed.

Assume that the system is man-made, meaning that the system state can freely be manipulated; the problem is that the inverse effect back from $\bar{x}$ to $\bar{u}$ typically is not optimized in the sense of emergy transfer. Still, it needs to be recognized that the property (3.36), and other observations therein, are general properties that always apply to the neocybernetic adaptation of ϕ in the form (3.13), so that the (visible) environmental signals are equalized. Assume that for physical reasons the feedback mapping is fixed F instead of adaptive φ.

If this inverse mapping F does not follow the "Hebbian learning" principle, so that (3.14) does *not* hold, does the whole cybernetic structure collapse? The answer to this question is *no*. Assuming that the feedback still can implement stabilization, the system in Fig. 3.3 will search a balance so that

$$\Phi^T F = I_n, \tag{3.50}$$

so that the feedforward and feedback mappings still have to be mutually orthogonal. Again, the feedback structure $\Delta u = F\bar{x}$ can be written as in (3.20); to make (3.50) hold, Φ again has to be given by (3.24), and it has to span the principal subspace of $\mathrm{E}\{\Delta u \Delta u^T\}$, because (3.29) still holds.

Even though the feedback structure in the system is fixed, the system properties remain essentially the same. The system cannot escape the subspace determined by F — but within that subspace, the model is optimized to tackle with observed variances. Even though the feedback matrix F perhaps cannot be affected, statistical properties of the signals can; after adaptation F spans the principal subspace of the converged environment as seen in signals Δu, making the originally non-ideal system ideal after all, in its own narrow world. The adaptation strategy does not allow trivial solutions, but excites the system by force. When the properties of the environment change, it starts reflecting the peculiarities of the system and its non-idealities — in this sense, it becomes questionable whether it is the system itself or the environment that implements the cybernetic adaptation.

It turns out that the latent variables can also be selected freely: Assume that there exists some $x' = Dx$ for some invertible mapping matrix D. Then, the original formulation $\bar{x} = q\,\mathrm{E}\{\bar{x}u^T\}\bar{u}$, when multiplied by D from the left, is identical with a new one, where only the variable x' is employed:

$$\bar{x}' = q\,\mathrm{E}\{\bar{x}'\bar{u}^T\}\,\bar{u}. \tag{3.51}$$

Utilizing these observations, the cybernetic studies can be applied for analysis of non-ideal real-life systems, where complete reciprocity of the data transfer structures does not *originally* hold. For example, in some cases these $\bar{x}_i'$ can be selected as the actual control signals acting on the system, as studied below.

Distributed controls

The above results make it possible to implement, for example, new kinds of sensor/actuator networks. In traditional agent systems, the issues of co-operation and shared "ontologies" are difficult; in the current setting, such problems become trivial: Each agent just tries to exploit the available resources in the environment. There is no need for negotiation as the interactions and feedbacks are implemented automatically through exhaustion of the resources. From the engineering point of view, it is nice that the goal of the agents — exhaustion of the variation in the environment — is parallel with the the goal of regulatory control (see [92]).

If the agents share the measurement information, transmitting the local measurements to the neighbors, the principal components oriented control of the environment is implemented after adaptation. If this assumption of complete information does not hold, the model becomes distorted: For example, if an agent only knows its own measurement, if there is no communication whatsoever among the agents, the operation of the control network becomes highly localized, even though there still is feedback through the environment.

As studied closer in the next chapter, the set of sensor/actors implements discretization of infinite-dimensional partial differential equations, the sensor/actuator nodes acting as discretization centers. Simultaneously the active participation of these nodes transforms the environment to fit the cybernetic structures. This control scheme can be applied, in principle, in chemical systems (the actuators adding chemicals if the measurements are low), or in thermal systems (the actuators heating the environment). An especially good application example is offered by mechanical systems, where the deformations and interactions are manifested practically delaylessly when some external forces are applied. In [92], examples of cybernetic "stiffening" of a steel plate are presented. This scheme can be applied also for design of mechanical structures, as shown below.

Design of mechanical structures

If the sensor/actor network is (virtually) extended over the whole mechanical construction, the network of controllers becomes more or less continuous; such settings can be studied, for example, in mechanical design systems that are equipped with finite element method (FEM), or, perhaps more appropriately, boundary element method solvers. Then one can apply the assumed forces onto the construction, calculate the deformations (or, more appropriately, the strains along the surface), and adapt the local controllers to oppose those deformations. After adaptation, there should be constant stiffness over the whole structure (see Fig. 3.6). The nice thing about this scheme is that the controls are manifested as increased stiffness, and the final "controls" can be implemented in terms of passive elements, simply adding extra layers of material in the locations of high experienced stress.

Today's design methods only take into account the maximum loads, and safety factors in specifications are needed to cope with unanticipated phenomena. Still, catastrophes take place every now and then — and typically the reason is *fatigue.* When the metal structures are under fluctuating tensions, the structures may break even though the specifications are never exceeded. Fractures are related to "gnawing". The cybernetic design approach — effectively damping and equalizing the vibrations — could offer new perspectives here.

The same idea of cybernetic designs could also be applied in frequency domain: At least in principle, (active) damping of vibrations can be implemented in this way. Similarly as in the static case of mechanical constructs, the system needs to be studied as a whole, as local damping actions can make damping efforts in neighboring nodes redundant, and iterative adaptation hopefully results in damping and equalization of vibrations. Here, the extension of the cybernetic framework to modeling of (discretized) functions is needed: The sensor/actor nodes host a family of input variables, these variables characterizing the measured energies at separate frequency bands in the power spectrum. Simpler implementation of vibration damping is reached if one concentrates on the velocities: Then the "information" being exhausted, or average of velocity squared, is proportional to the kinetic energy.

Optimization of parameters

The idea of cybernetic adaptation and constant stiffness against environmental disturbances can also be extended to large-scale industrial plants where there also is elasticity: A reasonably designed system can sustain environmental disturbations and other changes in a more or less robust way. Smoothly changing of parameters in the system (control parameters) or in the environment (set points, etc.) pushes the operating point of the plant in an elastic manner. Today, the low-level controls are typically poorly tuned, and separate control loops can have very different time constants, others being sluggish and other ones being faster. Uneven stiffness is manifested exactly in such heterogeneity between subsystems, and one can assume that cybernetic adaptation of the control parameters could make the subsystems better compatible.

In technical systems, however, there are domain-specific goals for evolution.

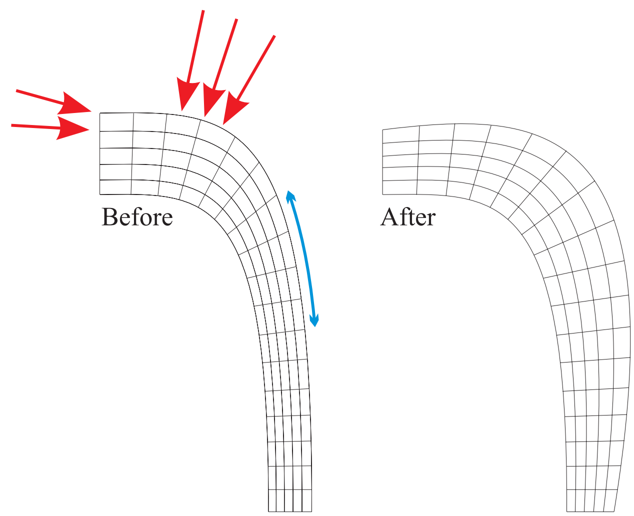


Figure 3.6: Iterative redesign of mechanical structures

One would not simply like to blindly adapt towards maximum emergy transfer, as assumed above, but one would like to maximize the match with the environment and the *intended* system functionalities. When the vector of system functionalities in x is predetermined by an external designer to contain some kind of quality measures, characterizing the "goodness" of operation, guided evolution is possible. The latent structure between u and x can technically be implemented in terms of not only correlations among variables in u (in the PCA style), but also in terms of cross-correlations between u and the intended x (in the PLS or CCR style, for example — see [42]). When the design parameters in u are seen as variables on the slower time scale, evolution towards better parameter values implementing higher values of $\bar{x}$ can locally be seen as pressing the elastic system into a desired direction along the determined axes of "quality freedom" (see [92], Report 139: "Process Performance Optimization Using Iterative Regression Tuning").

3.4 Towards *complex* complex systems

How are the above abstract assumptions about evolution related to real-life observations? Indeed, it seems that *increase of stiffness,* or *hyperplasia,* is the key behavior in natural adaptation processes. For example, skin becomes thicker if it is burdened, and a muscle becomes stronger if it is used (reactions of the neural system to signal activation are discussed in detail in chapter 7). Similarly, companies invest money and employ new staff if there is very much activity. This kind of trivial-looking behaviors, when boosted with self-regulation and self-adaptation, result in global-level system properties that can be described in terms of principal components. Because of the properties of PCA, adaptation in the assumed form maximally compensates the external disturbances.

This far, simple systems have been studied. The key observation is that, when seen at the correct level of abstraction, "all" complex reasonable systems are elastic. Elasticity offers tools to attack really complex, formless systems. From

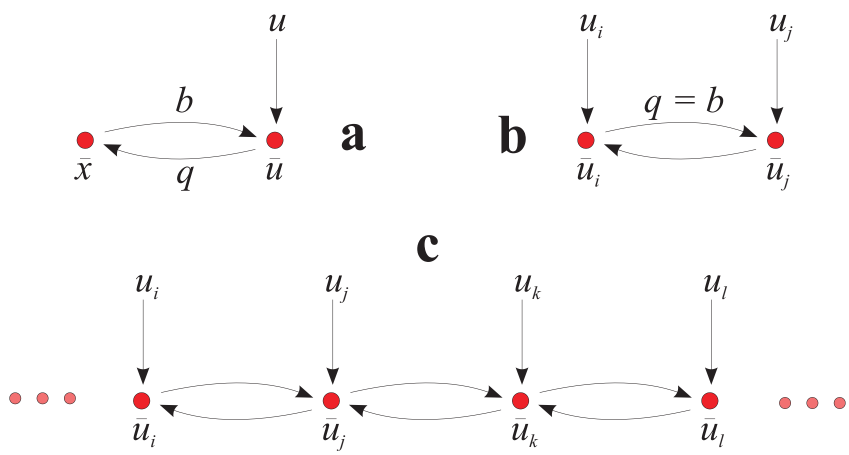


Figure 3.7: Extending the cybernetic framework: The studied cybernetic system structure (a) is symmetric what comes to the interactions, and the roles of the system and the environment can be shared (b) — and, finally, the environmental variables can be physically distributed (c)

now on, no accurate mathematical analyses are available any more: One just has to trust in the strong modeling principles, and intuition. These elasticity ideas are closer studied in the subsequent chapter.

As a brief introduction to extensions of elasticity considerations, look at Fig. 3.7. It turns out that there can exist a wealth of neighboring systems that are more or less tightly connected together; from the point of view of a single subsystem, the neighbors together constitute the environment. As it is various neighbors that see the same system as their environment, the coupling factors q in different subsystems must become identical as they see the same level of variation in their environments. As seen from outside, it is only the coupling coefficients that remain, determining the dynamic properties of the system. As the number of neighbors grows, dynamic transitions become diffusion processes among differential elements. In any case, the local adaptation as presented before still gives consistent results. In physical systems the interactions are concrete, but they need not be — it is all about information transfer. Interchange of the roles of the system and the environment is studied in more detail in the next chapter.

As a conclusion of this chapter, it can be observed that within the neocybernetic framework, local learning has globally meaningful results. As seen from functional point of view, new interpretations for cybernetic systems are available:

- **First-order cybernetic system** finds balance under external pressures, pressures being compensated by internal tensions. Any existing (complex enough) interacting system that can maintain its integrity in a changing environment is cybernetic in this sense. First-order cybernetic system momentarily implements *minimum (observed) deformation emergy in the system.*

- **Second-order cybernetic system** adapts the internal structures to better match the observed environmental pressures, towards maximum experienced stiffness. Any existing (competing) interacting system that has survived in evolution finally is cybernetic in this sense. Second-order cybernetic system additionally implements *minimum average observed deformation emergy in the system.*

- **Higher-order cybernetic system** adapts the external structures of the system to better match the observed environmental structures by adjusting the impedances. Evolutionarily optimal environment, or system of systems, assumedly only contains higher-order cybernetic systems. Higher-order cybernetic system implements *maximum average transfer of emergy through the environment.*

Level 4

Systems of Populations as *Symbiosis of Agents*

Studies of systemic biology must not be restricted onto the cellular level — after all, another class of truly challenging biological phenomena concern *populations.* One needs systemic means to understand the superposition of individual actions. Today's understanding of *ecologies* is surveyed, for example, in [48].

Again, abstracting away details gives statistical models capturing the population properties in the large. There exists a wealth of first-principles models that are tailored to explicitly explain certain ecology-level phenomena (for example, see [57]). However, applying the straightforward modeling principles and letting them cumulate, the originally simple (nonlinear) formulas soon become very complicated, and the value of the models becomes compromised. One should keep in mind that it is individuals only that exist, giving rise to the emergent properties: It can be claimed that without taking the actors into account gives incorrect intuitions. Not whatever simplifications are appropriate when modeling populations — again, it is the neocybernetic principles that should be applied. This time the model structures already exist, and the "off-the-shelf" models can be directly exploited.

Neocybernetics offers tools to capture the population-level global properties that emerge from local interactions. It has been said that ecology is systemic biology; now one can reach towards *systemic ecology.* Ecosystems with the same underlying principles also exist outside the boundaries of conventional biological thinking, and examples are presented from other fields to perhaps reach interesting cross-fertilization among disciplines.

4.1 Extending from a domain to another

In the previous chapters, concrete examples were studied in a bottom-up fashion, finding holistic views characterizing the system-level phenomena. Now when the concepts have intuitive substance and content, it is possible to continue the studies in other domains, exploiting the same understanding, without any more

Figure 4.1: It is the neighboring systems (agents) that *are* the observed environment (graphics by Maurits Escher)

going into details. There now exists understanding about the crucial nature of complex feedback structures within the system: One does not need to know exactly how they are implemented — but to maintain the system integrity *they just have to exist there.*

4.1.1 Environment seen as neighbors

How to define a *system?* It has been said that the most complete definition can only be such that

> 'system" is a *system.*

This "definition" employs our intuitive understanding — what can be seen as a relevant entity, functionally consistent host of dynamic attractors, can be seen as a system of its own. It can be claimed that all other definitions of system are too narrow, and would not cover all aspects of the idea. However, such heuristics is a challenge to the traditional systems thinking.

Traditionally when analyzing systems, it is the system boundary that is perhaps the most important thing to characterize: The boundaries separate the "inside" (the system) from the "outside" (non-system). According to the selection of these boundaries, variables coming from outside are seen as input signals. When doing neocybernetic modeling, however, even the basic conventions are changed: The systems and their boundaries become relative, dependent of the point of view. Now *there is no separate environment* — when the feedbacks are seen as an integral part of cybernetic systems, environment becomes an essential part of the whole, the system properties being determined together by the environment and the actual system. As John Donne almost said, "no system is an island".

The purpose here is to extend the studies from the realm of inner-system phenomena onto the environment. In principle, one is stepping from the (assumedly) known into the unknown; however, it turns out that it is the same ideas that hold inside and outside. One could say that *somebody's environment is somebody else* — it is other (more or less) similar systems that are found outside. The neocybernetic model formulas are reciprocal and they can be inverted quite formally without changing their structure — the “inside” becoming the “outside”, and *vice versa.* One only turns from studying a single system towards studying a set of such individuals. The system’s “inputs” are (mainly) neighboring systems’ state variables, and *vice versa.* Indeed, when the internal structure of the combination of systems becomes more complicated containing more variables, the *actual* independent inputs coming from the outer world are less visible than the “internal inputs” (see Fig. 4.1).

The uniqueness of the system boundaries becomes challenged also in other ways. Traditionally it is hierarchies of systems that are used to structure complex domains, but now one cannot determine a hierarchy of subsystems in an unambiguous way — the appropriate structure depends on the point of view, depending on the level of accuracy, and the selection of variables. A single subsystem can concentrate on a single functionality, or it can take care of more functionalities — that is, n can be 1 or higher — no matter how the boundaries are selected, self-organization reconstructs the functional structure according to the signal properties, the same (linear) principles operate on all scales.

Assume that each cell stands for a single functionality only, the functionalities being different for all cells. Further assume that diffusion distributes the signals (chemical concentrations) evenly among the cells. The system of subsystems finds its balance, the subsystems exhausting each other’s “waste products”. This is the simple basic scenario for explaining *symbiosis* — relevant functionalities are distributed among localized actors (in this case cells). It needs to be recognized that there is no “negotiation” or higher level operation control needed: All cells just optimize their behaviors in their very local environment (and, indeed, the cells themselves do not even know they are carrying out some optimization).

More sophisticated structures of symbiosis are readily imagined. Assuming that each cell is alone responsible of only differential effects, it is the steady state values $\bar{x}_i$ that reveal the proportions of different functionalities, or cell types, that are needed to fulfill the environmental needs. When there exist individual cells sharing the same functionality, the capacity limitations are compensated by the high number of identical subsystems. The balance ratios of the numbers of representatives for different types are determined by the balance values $\bar{x}$. It is all about dynamic balance pursuit again.

The symbiotic “systems of systems” are characteristic also to more complex domains: From the level of individual cells one can get to the level of tissues and organs; from the level of individual organisms one can get to the level of populations; and from the level of species one can get to the level of ecosystems — what is more, the organisms and their systems can be physical or abstract. This kind of characterization is possible in principle. As the systems become more complicated, however, the signals and interactions become more abstract, and no conclusive models or predictions can be made. When escaping the immediate chemical domain, there is more freedom to construct the systems. Cognitive systems (symbiosis of neurons) will be studied in chapter 7, and the role of different views of the available information is studied in the chapters 5 and 6. Some ideas concerning symbiotic systems are presented below.

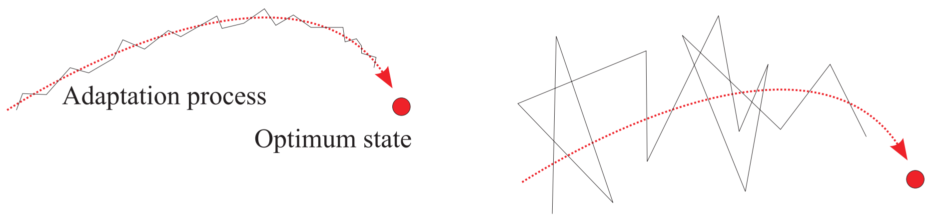


Figure 4.2: The adaptation strategies and dynamics can be very different in different kinds of systems — but the final state, the hypothetical optimum, is the same for all

4.1.2 From individuals to a population

When extending the analyses from symbiosis of cells to symbiosis of populations, in principle, it is easy to reinterpret the symbols: The vector $\bar{x}$ represents *population sizes,* $\bar{u}$ is the vector of *available resources* (different types of food and other living conditions), and matrices A and B contain the *interaction factors* (competition) among populations. The columns in ϕ can be called *forage profiles,* exploitation conventions corresponding to the populations. But is this more than renaming? Are the very different systems really analogous? And, after all — is there universality among complex systems?

When abandoning the familiar domains, all points of support seem to be lost: If studying distinct organisms, the chemical cues, for example, can be completely secondary when the interactions become implemented. As the environment is seen as consisting of individual subsystems, it is no more mere signals that can be measured in the environment; the higher one gets in the hierarchy of systems, the more the inputs become more and more abstract *functions,* as it is the functionalities $\bar{x}_i$ that are used as inputs $\bar{u}_j$. It is assumed that such functions offered by the environment (or *need* of functions as requested by the environment) can somehow be quantified. In any case, it is the basic intuitions that remain: There must again exists the same kinds of underlying principles to make the emergence of organization from non-controlled local behaviors possible — there must be common pursuit for survival shared by all agents. But there can exist many alternative ways to survive in the environment — why should one favor one specific model structure?

The key point is that *following the neocybernetic model, there is evolutionary advantage.* It turns out that *optimality in terms of resource usage is reached,* meaning that surviving, successfully competing natural populations assumedly must have adopted this strategy. As soon as the coupling coefficient q in a system reaches the threshold in (as discussed in chapter 3), there exists a clear evolutionary gradient visible towards the optimum. This all is mathematically very simple, and as long as the neocybernetic strategy is the only known route to self-organized principal subspace analysis, it can be claimed that there exist no competing theories. In the long run, it is the models that implement the PCS model that can best be matched against variations in the resources $\bar{u}$ (in terms of quadratic variation criteria), resulting in most efficient exploitation of the resources. And populations with optimal strategies outperform others in terms of biomass and more probable survival.

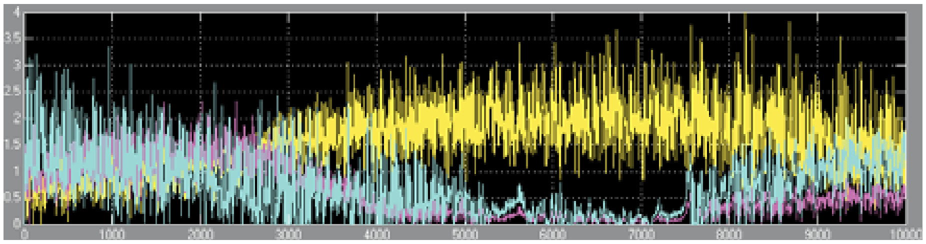


Figure 4.3: A typical simulation illustrating the behaviors of three (hypothetical) competing species (in arbitrary units). The environmental conditions, or resources in the vector u, are random but have a certain statistical distribution, and the populations $\bar{x}$ are assumed to instantaneously follow the changes. Population sizes below zero are simply zeroed during adaptation — finally this results in emergence of populations remaining always positive-valued

In the framework of resources and their exhaustion, "power transfer" between systems can be made concrete: It is assumed that the product of consumed resource and produced activity has the dimension of power, as being reflected in reproduction capability. — Of course, the whole theory collapses if there are no variations in the signals, if studying static environments. In such a case no strategy can be claimed to outperform the others; these issues are studied on Level 5.

It is perhaps hard to believe that the very nonlinear genetic mutations and accommodation processes, etc., would have anything in common with the cellular adaptation details. How could the same model apply? The key observation here is that it is, again, only the dynamic equilibria that are studied, not the all possible routes there. Whereas the adaptation processes can be very complicated and varied, the final emergent optimum can be unique in terms of tensions (see Fig. 4.2). When concentrating on the balance only, it is also the dimensionality of the problem that goes down, making the optimization process feasible. And, remembering the previous chapter, it is the "interfaces", common variables between interacting systems only that count.

4.1.3 Properties of a cybernetic population

Traditionally, ecological models concentrate only on a single species or interactions between two species (for example, see [79]). Larger models try to characterize the *niches,* implementing explicit *forage profiles* that describe the resource specifications for each species [75]. However, such models for complete ecologies need careful tuning; evolutionary strategies typically become unstable, meaning that most of the species become extinct, only some of them prospering and exhausting all resources.

When applying the neocybernetic model, ecosystem simulations remain stable even though the dynamics looks "naturally chaotic": There exists unforced dynamics in different time scales (see Fig. 4.3). Adaptation in the system is based on cybernetic evolution — there is vivid dynamics, but no explosions take place. Rapid stochastic variations in the population are followed by slow long-term behaviors. No fine tuning is needed: If there is enough variation in the resources, after system-wide adaptation a balance is found where there is a "niche" for every species. The niches are characterized by the principal subspace dimensions, the forage profiles ϕ_i mapping from the prevailing resource vector $\bar{u}$ to the balance population $\bar{x}_i$. The roles of the species cannot be predicted, only the subspace that is spanned by all of them together is determined by the environment. The key observations concerning the neocybernetic model properties can be summarized:

- **Robustness.** In nature, no catastrophic effects typically take place; even key species are substituted if they become extinct, after a somewhat turbulent period. Using the neocybernetic model, this can also be explained in terms of the principal subspace: If the profiles are almost orthogonal, in the spirit of PCA, changes in some of the latent variables are independent of each other, and disturbances do not cumulate. Also because of the principal subspace, sensitivity towards random variations that are not supported by the long-term signal properties are suppressed.

- **Biodiversity.** In nature, there are many competing species, none of them becoming extinct; modeling this phenomenon seems to be extremely difficult (see [89]). Now, again, this results from the principal subspace nature of the model: As long as there exist various degrees of freedom in input, there is reason for different populations. Within species, this also explains why in balance there exists variation within populations as the lesser principal components also exist (compare to the *Hardy-Weinberg law:* "In a large, random-mating population, the proportions of genes tend to remain constant from generation to generation").

4.1.4 "Complete-information ecosystems"

Regardless of the uniqueness assumption concerning the optimum state, practical manifestations of the underlying dynamic balances vary a lot: What kind of populations will exist is not only dependent of the environment, but it also depends on the physical constraints. In nonlinear systems it is not only the final balance that is relevant — the route towards the optimum makes a difference, as the process can end in local minima. As will be discussed on Levels 5 and 6, it is the availability of information that makes a difference, and nonlinearities can often be interpreted in terms of different kinds of information blockages. For example, how the information theoretically motivated resource variation coverage is carried out in an ecosystem, depends on what kind of species are available — information cannot cross the species-wise genetic pools at the same rate. All these blockages together give rise to non-logical outcomes even in the equilibrium. During this chapter, however, complete availability of information is assumed. In practice, this means free mobility and information transfer among the signal carriers within a specific phenosphere.

It turns out that from the simplest chemical levels, it is easiest to skip all the intermediate levels (tissues, organs) directly to the most challenging levels, to the least structured ones, consisting of populations of more or less "intelligent agents", where complete information exploitation can be assumed. The intermediate levels necessitate more explicit structures, or *differentiation* among populations, and nonlinearities are necessary, as will be studied on Level 6.

The knowledge is also assumed to be shared equally by all actors in the system. It is assumed that all members of the populations can recognize all resources, and their weighting of different types of resources is similar. Whether this can be assumed or not in natural populations is to remain an open question here — but, again, when seeing ecosystems in a wider perspective, fruitful analyses can be continued.

It turns out that ecosystems need not be ecological — they can also be *economical.* There are attempts to apply holistic thinking to economy (for example, see [76]) — however, those models are constructed in the traditional way, bottom-up, trying to capture the system's properties in a collection of its parts, and a wider view is needed.

The above cybernetic discussions can somewhat directly be applied to market economy: Companies stand for species, and variables $\bar{x}_i$, or "population sizes", are company turnovers; input $\bar{u}_j$ is the available "benefit" in the product group, and the vector ϕ_i characterizing the company contains its production profile (other interpretations for the symbols are also possible). Individual humans are only "signal carriers", like ants in an ant colony. Strategies dictate the company-wise (or less wise) adaptation styles, as being manifested in economic decisions involving recruitment policy, investments, etc. Adaptation in a company is very nonlinear and non-continuous – however, if the company is to survive in the competition, the stochastic processes have to be more or less consistent in the long run, resulting in the balance determined by the environment. For example, the growing system stiffness becomes implemented in a natural way, new workers being employed if there is need for them, if there is work overload. From the point of view of the whole system it is statistically irrelevant how resources are distributed among the companies – however, for a single company, the details make a big difference: An individual company may prosper or suffer, or get extinct. Yet, in the case of bankruptcy, the system soon fills the niche with others companies.

What makes such an extension to a still more complex domain motivated, is the fact that *quantification of resources and efforts becomes easy* in abstract enough systems. In an economy, the universal measurement stick for "benefit" is *money.* All variables can be made structureless and dimensionless, all things become commeasurable when they are put on the money axis. When the role of money is generally accepted, and when the prices have been agreed upon, the cybernetic system can become more efficient, streamlined, and transparent. "Everything has its price" is the truth in an efficient economy; it is irrelevant whether or not this is ethically sustainable. One has to forget about morals: There is no "good" or "bad" in nature.

So, in principle, market economy operates like an ecosystem, and numerical analyses and simulations can be carried out. However, there also exist features in there that motivate a still wider scope. In different kinds of *memetic systems* one does not necessarily speak of money, but it is the same types of evaluations and assessments of alternatives that are being carried out, balancing between different kinds of visions of the reality. Such memetic systems include, for exam-

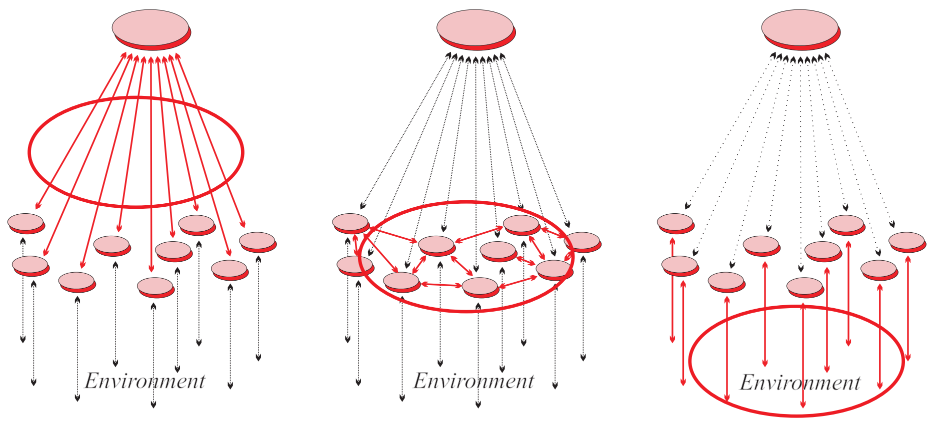


Figure 4.4: Communication and coordination among agents. The new view of differs not only from the centralized approach (on the left), but also from the traditional distribution, where the coordination is based on explicit communication (in the middle): Now the emphasis is exclusively on the environment (on the right)

ple, *scientific, social,* and *political* arenas. When politicians speak of "values", this should actually be taken in quite concrete terms: Everything can be traded, only the prices vary. Tensions in the systems are caused by contradicting aspirations. To understand such *constructivistic systems,* the role of the humans needs to be elaborated on.

4.2 Agent systems

There is a huge conceptual leap from a concrete domain — chemical signals, for example — to the domain of more or less *intentional* actors like animals or humans. Because of the intuitive differences, it is perhaps appropriate to speak of *agents,* when the signal carriers are functionally such independent entities with more or less "free will". The question is, again, very much about appropriate connotations. The discussions necessarily become very deep (for example, see [30] and [7].

4.2.1 Humans as agents

The agent paradigm has become popular as a framework for studies on distributed intelligent systems [67]. Today it is mostly about "software agents": The agents have been explicitly programmed to behave appropriately without centralized control [51]. To implement cooperation among the agents, complicated communication protocols and common *ontologies* are needed, and it is difficult to implement some adaptive behaviors in such systems. Now, on the other hand, one has trivial data-based ontologies and semantics: Everything important in data is buried simply in correlations between signals. There is no need for communication strategies, because the only interaction takes place in the form of feedback through the environment; indeed, there is a new view of seeing the structure of distributed systems (see Fig.4.4).

When extending the studies in the previous chapter to more or less intelligent agents, when claiming that systems consisting of humans, too, implement similar emergent functionalities, one has to motivate why the simple model structures still are applicable. For example, what are the "mental degrees of freedom"? As shown in chapter 7, neuronal elasticity in the realm of signals can easily be motivated, but how about the higher level, the visible level of mental functions? Indeed, at least there are some qualitative intuitions that are supported: In a new situation the behavioral spectrum is wider for a novice than for an expert. Through cumulating life experience the "mental stiffness" increases as the process of automatization is manifested in all types of behaviors. Because of the difficulty of quantifying abstract phenomena, nothing very concrete can be said here — however, as studied in chapter 5, the variables in one's subjective world become the measures that characterize the objective world, too.

How about the learning strategy Hebbian-style — is that obeyed by humans? There are two opposite mechanisms that are needed: Further adaptation towards the resources in the case of intense activation, when doing the "right things", and reallocation of efforts in the case of deprivation, when doing the "wrong" ones. And, indeed, it can be claimed that there exist the two classes of basic low-level mechanisms that seem to be wired in human brains — and, at least to some extent, also in less sophisticated animals:

1. **Motivation strategies.** When one invests effort in reaching some resources, successfully managing in doing that, this behavioral pattern is typically strengthened. In other words: When one's activity $\bar{x}_i$ is rewarded implicitly in terms of resources in u, one's behavioral profile is adapted towards such practice, so that there will be more activity in that direction. Clearly, this is very well in line with the assumed cybernetic learning strategy. For humans, the resources and rewards need not be concrete — the "resources" in u can be different kinds of *possibilities* available in the environment, and the reward can also be provided in terms of *encouragement* by some external critic. On the other hand, similar (but less abstract) learning behaviors have been observed also in animals that can modify their forage habits according to the available prey; in lower life forms, the conditioned reflexes are a manifestation of the same non-genetic accommodation processes.

2. **Compensation strategies.** What happens, on the other hand, if an agent never succeeds in its strivings, always being turned down? Among lower-level organisms, when the competition concerns food and other physical resources, this typically results in the organism gradually fading away — however, for intelligent agents the results need not be so acute: When the resources are interpreted as possibilities, there exist other options. The psychological concept *compensation* means the mental reaction against disappointments (among other protective mechanisms), where one claims that "it was not what I wanted really". In a way, it is only self-deception — but it is very real in one's subjective world. The "local reality" will be re-evaluated, the unattainable resources are understated; in technical terms, this means that one's weightings of the variables are redefined.

> This may change one's orientations altogether: Hopefully one can finally exploit and further develop one's personal capacities, when the domain of one's special talent is found.

When studying both of the above mental processes, and the corresponding behavioral adaptations, the creativity of a human seems to be endless. There always exist new ways to see the world and its possibilities, and new ways to better manage are invented. When a human implements system adaptations, wider views and understanding is available: When there exist various separate cybernetic systems coexisting in the same mind, cross-fertilization of ideas becomes possible.

Animals can actively not only change the weightings of the variables, but they can also change the contents of them — moving to another environment, for example — but human's flexible mind makes it possible to change not only the values but the variables themselves.

Introducing new variables means that the structure of the system changes. Structural changes can be implemented in a human mind in a very efficient way. Intelligence makes it possible for an agent to escape outside the existing constraints of the prevailing ways of seeing the world. However, free will is a fallacy — with a twist: Even though quite original behaviors are possible in principle, not obeying the basic cybernetic principles is not seen as personality — such destructive behavior would generally be regarded plain insanity.

4.2.2 Intelligent organizations

When seeing in the wider perspective, the above selfish strategies result in merciless competition. In the social animals, there seem to exist mechanisms to easier find the sustainable society-level dynamic balances: Some primates share a simple hard-wired social strategy that could be called *monkeying,* perhaps the simplest form of co-operation. It is typical also for humans to mimic behaviors and follow the leaders, making it easier for organization in low-level societies to emerge. But humans are not bound to the hard-wired behavioral patterns: Cultural evolution has bypassed the biological one.

There are some functionalities that are necessary for a cybernetic agent: One needs *sensing, inference,* and *memory,* or, in concrete terms, measurement and analysis of signals, recognition of correlation structures, and storage of these structures. If the agents have more capacity, it is possible, for example, to implement different kinds of optimization strategies. Whereas the simplest agent only tries to survive, not taking other agents into account, blindly exploiting the resources it can see in its extremely narrow local world, a slightly more sophisticated agent can see its environment in a wider perspective: It can see the actions of its neighbors, and it can actively start avoiding competition. A more intelligent agent not only tries to go for resources according to (3.13), but it also escapes neighbors, predicting the interactions and taking the feedbacks into account beforehand, changing the "Hebbian learning strategy" into "Hebbian/anti-Hebbian strategy" [92]. Such higher-level strategies make it possible to reach emergence of higher-level structures (principal subspaces, etc.) also in rigid, non-flexible environments, where feedbacks will not become implemented through the environment (this is the case, for example, if the population exploiting the environment is negligible as compared to the amount of resources). And if an agent can see still wider horizons, it is possible to optimize even further, reaching towards cybernetic optimization, balancing among a network of neighbors; such balancing becomes easier if there is cooperation, and if the agents can somehow negotiate.

Communication, or "negotiation" among agents in its simplest form is mere signal transfer: It is not necessarily a purposeful act, as distribution of the physical chemical levels can be seen as transmission of variable values as well. In lower animals and especially in insects, chemicals play a central role in communication not only within a single organism but also among them. In the world of smells and scents, low concentrations of thousands of chemicals can be detected, each of them augmenting the vector of available cybernetic variables, and it is pheromone transfer among the signal carriers that guides the construction of ant hills, for example. In higher animals and especially among humans, on the other hand, more complex coding of messages takes place. It is the quality and extent of communication among humans that determines the limits of *learning organizations* as discussed in [71] — issues concerning "information bandwidth" are discussed in chapter 5. Because the principles are (assumedly) identical on different layers of cybernetic systems, the functions of a single human can be extended in an organization of several humans assuming that information transfer can be provided in a seamless way. The possibilities of the organization to construct networked intelligence is dependent of the agents' abilities to understand each other; in this sense, perhaps it is the EQ, or *emotional intelligence quotient,* that determines the capacity of the higher-level intelligence. It is all about information transfer — systems can become one if the communication among them is tight enough.

But if an agent is (too) intelligent, it probably uses its wits to choose itself how to utilize the information. To avoid anarchy and to reach evolutionary advantage on the society level, not whatever game theoretic optimization in the intelligent agents can be tolerated as a general rule, however. To reach emergence of higher-level patterns in the global system, it is beneficial if the agents share some common code. There are no genetic chains, but there are cultural ones. After all, the agents need to be *humble,* obeying some *categorical imperatives* that somehow persuade or press the agents to think about the common good. It is not a coincidence that in all prospering human cultures there have been some religion dictating the roles for the individuals. However, the moral codes can be implemented not only through religion, but also through philosophies or through secular legislation — somehow expressing the Kantian idea "only act in such a way that wouldn't ruin the system if *everybody* acted that way". And, today, the modern imperatives guiding towards the "economical optimum" are implemented in terms of *fashions,* etc. Culture and social codes help in finding one's niche without wasting ones energy in vain, kicking against the pricks. There is a niche for only one clown in the classroom.

There are humans with varying properties. In an intelligent organization this is taken into account, and the tasks and workloads are organized according to individual abilities. A team needs its organizers, "mood makers", etc.; the traditional line production style optimization is not cybernetically robust. As the humans learn their jobs still better, adapting in their environments, counteracting the local tensions, finally finding their niches, the operation of the organization becomes streamlined. In the cybernetic framework the evolution theory can be extended to human cultures without having to employ the cruel ideas of "social Darwinism".

One can also draw conclusions concerning wider organizations. The "Adam Smith Type" capitalist economy is efficient, explicitly implementing the survival of the fittest, but a designed welfare state based on global-level optimization directly implementing the global equilibrium can be even *more efficient* — *assuming* that the underlying model of the society is correct. The laws, etc., should be determined based on correct predictions of the system dynamics and inertia. The less one can afford experimenting and iteration, the more accurately one should capture the balance already in design. Intelligent, more cultivated strategies make it possible to avoid needless competition — struggling for life, or *suffering* in general, as studied in the Eastern philosophies, and also by Western philosophers like Arthur Schopenhauer.

To truly understand the universality of cybernetic thinking, one can extend the studies from *autocybernetic* to *allocybernetic* systems: This far, the actors implementing the functionalities have been themselves part of the system; now, however, it is assumed that the actors operate in *another phenosphere.* Still, after the variables are evaluated and the adaptation is carried out, it is the general cybernetic principles that determine the domain area structure regardless of the intentions of the actor.

4.2.3 Constructivistic systems

The extended intelligence in agents can be exploited in different ways. It is all the capabilities that can be enhanced: New variables can be made available, more storage can be allocated, and structures of correlation can be defined in new ways; "creativity" is the key to clever combinations of new capacities. There can exist various overlapping optimization processes with different sets of variables taking place simultaneously in one agent, opening up a wealth of cybernetic, fractally overlapping systems. A human agent with his/her magnificent mental capacities can operate on phenospheres that this agent is not physically part of. There exists no exact borderline between autocybernetic and allocybernetic systems — the larger the system is, the more the system follows its own dynamics, and the human can only look aside; or, indeed, the human changes to a mere signal carrier.

When the agents are human beings, perhaps the most characteristic allocybernetic systems are located in *infosphere* or *ideasphere.* The "idea atoms" are called *memes,* and they are the "genes of the infosphere", being the building blocks to be appropriately combined [20]. The term "meme", referring to some concrete idea, is not always quite appropriate; typically, needs and desires are not necessarily ever explicated. The human mind supplies for the platform for the memes to operate: Humans collectively host the memetic systems, acting as the signal carriers, and providing the machinery for emergy processing. Even though constructivistic systems are explicitly designed by humans, after the memetic platform has been consistently instantiated, so that reasonable balances among "memetic tensions" can be found, it will start following its own dynamics, the memes trying to prosper among competitors, assumedly obeying the cybernetic principles. The final success criterion is the match with the environment, memes together explaining the observation patterns. To constitute a cybernetic system, there needs to emerge a dynamic balance, meaning that there needs to exist a competition of counteracting memes finding equilibrium. The human's contribution to the memetic dynamics is that he/she defines the environment for the memes, coupling the variables to the real world, choosing the relevant observations (variables) and interpretations for them, also supplying for their evaluation (weighting of the variables). The emerging structures depend on how the world is seen — how the abstract and non-concrete phenomena can be quantified, is further studied in Sec. 4.3.

Emergent structures in infosphere are, for example, theories or paradigms (when talking about science), or "isms" (in politics). Study some examples of such complex domains — there may be some cybernetic intuitions available.

Society and politics

In a social system, and specially in politics, the "variables" are those issues, resources, possibilities, and needs, that are being discussed and debated. The intuitive common goal is to reach a systemic balance state where there are no more unjust evils. Of course, visions of the utopia differ as the weighting of different issues differs among citizens in a pluralistic society. The views of ideal society, or the agenda of aspirations, becomes manifested in the programs of parties and profiles of candidates in elections. How to measure and quantify the state of the complex social system, then? In the political arena, the contradicting aspirations are made explicitly quantifiable through a kind of Analytic Hierarchy Process [68], where alternatives are given to voters to choose from: Different opinions become quantified in elections. Popularities of parties (number of votes in $\bar{x}$) can be assumed to reflect the vector of needs $\bar{u}$ in the society — and, in the democratic system, this popularity is reflected in the capacity of implementing the party visions. In a way, the role of candidates (and parties) is to act as *probes* to identify the infinite complexity of the society (see Sec. 4.3).

Why democracy seems to prosper even though it is less efficient than a (Platonian) dictatorship, why is democracy typically restored even after turmoil periods? Assuming that there is complete information available in the society, democracy represents the *most cybernetic political system,* giving the maximum information from the bottom level to the top, thus keeping such a system maximally up-to-date. Parties determine profiles of opinions; party popularity (number of votes $\bar{x}_i$ corresponding to the party profile ϕ_i) reflects specific needs $\bar{u}_j$ in the society, and this voter support is reflected in possibilities of implementing party visions. When representatives are selected, not all decisions need to be brought to the ground level. Is the current system the best possible, then?

> In the era of enhanced information technologies, more sophisticated voting practices could be employed, for example. The current voting scheme is too coarse to reveal the nuances in opinions. Why not allow a *spectrum* of votes, so that votes could be distributed among candidates, the total weight of one's votes still equaling 1? Today,

> each party has to be the voter's only choice, making it necessary to become a "general-purpose party". In a long run, this results in a democracy that cannot respond to changes: When all parties become identical, no structure among the parties emerges any more. In principle, in a fully cybernetic society, the parties should span the principal subspace of existing aspirations; now, when only the mainstream averages are followed, the developments in the society become more like random search process. What is more, different kinds of thresholds, etc., jeopardize the linearity of this model of the society.

In any case, each tension (or aspiration) has to be finally compensated by counter-tensions — otherwise, the system becomes pressed endlessly, and the system collapses.

Seeing the politics as a cybernetic system perhaps makes it possible to understand and react to the pathological developments. For example, today it is no more possible to restrict the "variables" to those issues that one would like. In the postmodern society there seem to exist no real acute problems, and criteria are in a process of change: Politics is becoming entertainment, debates become "true television", where substance is substituted with appearance. Or, putting it more philosophically: The cultural patterns emerging from the human values and aspirations reflect the *Weltgeist* in the spirit of Hegel.

Scientific communities

Similarly, in scientific research there are complex domains to be explained: Together the theories span the space of observations so that a reasonable balance between models and reality is reached. In principle, the most important criterion for a good theory is the match with reality — but the reality can be seen in different ways, or the relevance of different phenomena can be assessed in different ways. In natural sciences, the external world really exists and there are concrete measurements available, but also there, it is the interpretations and internal dynamics of the scientific community that plays a central role. In the postmodern era of "ironic sciences" (as discussed in [40]), the similarities among branches of scientific work are becoming more and more evident, and *consilience* seems appropriate [90].

The central challenge in all scientific work is to define the variables and their weightings: What is relevant, and how important it is. These issues are settled when the framework is fixed: This framework can be identified with the *paradigm* in the Kuhnian structure of sciences [49]. Within the paradigm, there are theories, or scientific memes, competing for popularity, individual researchers just acting as information carriers. Such paradigms are rather stable attractors — as long as they can sufficiently address the real-life challenges. But after new sets of theories clearly outperforms the previous ones, the paradigm shift can be abrupt.

Determining what is "good science" is a specially challenging task. By definition, science should tackle with something that is unknown and unstructured, so that no *a priori* weighting can be reasonably defined. Instead of the subtle contextual criteria, better quantifiable bureaucratic guidance is becoming more and more dominant: It is easy to define numeric measures — like number of publications and amount of publicity, etc. Todays answer, common to all branches of scientific work, is to trivialize the problems, inflating the strictly scientific criteria. Also the criteria based on peer-reviews are problematic: When goodness of research is defined in terms of match with the scientific community, a scientific paradigm becomes a self-sustained entity. Science is what scientists do — as studied below, reality is molded by the actors — or, indeed, reality is *created* by them.

Yet, however long the wrong tracks are, the Darwinian dynamics in science is extremely efficient and cybernetic. When some scientific branch is most active, new interesting facts being detected, it also is most adaptable: There are bright minds and financing available, making adaptation fastest in the directions of maximum benefit. On the other hand, nobody feels pity of the losing theories, such researchers having to search for new directions, resulting in structural rearrangements in that field. This all is perfectly in line with the Hebbian-type learning. As contrasted with economic environments, there is a clear difference: In science the idea is to "change the behaviors when the times are good", but in companies the principle is to "not fix if it still works" — adaptation taking place only in bad times!

Case: Stock markets

As an example, study a better quantifiable domain field that is explicitly constructed but whose dynamics is still beyond control, and even beyond comprehension: *Stock market* was originally created for balancing the imbalances in economics, but today it seems to have escaped the controls, following its own chaotic dynamics. It seems that such a domain field offers a possibility of more or less immediate application of new thinking; it has become an independent cybernetic entity itself, pursuing balance but being vulnerable to catastrophes (see chapter 5). Contrary to the claims, the behaviors in the market cannot be reduced to the economical fundaments. The stock market is a prototypical example of a yielding elastic systems: As the demand rises, the price goes up until the balance is found. This balancing is very fast, and also adaptation of the system is fast, as money is transferred in principle without delay and the model structures exist only in the form of expertise in the analysts' brains. Stock market truly is an extreme example of maximum exploitation of information on the edge of understanding (chapter 5).

The underlying dependencies among the exchange rates are not known — but they need not be known. Indeed, because of the fast information exploitation, the stock market can be seen as being in a dynamic equilibrium where opposite drifts balance each other; what is more, as the agents all the time try to maximize their profits, balances are being tested all the time, and there is maximum regeneration of information. The balance indeed changes to generation of excitation, so that there is inherent drive towards the edge of chaos.

In principle, there is the underlying real world that determines the market prices — however, the dynamics is detached from the underlying realm, behaviors being based on internal tensions. It is like it is in *cognitive systems* — the

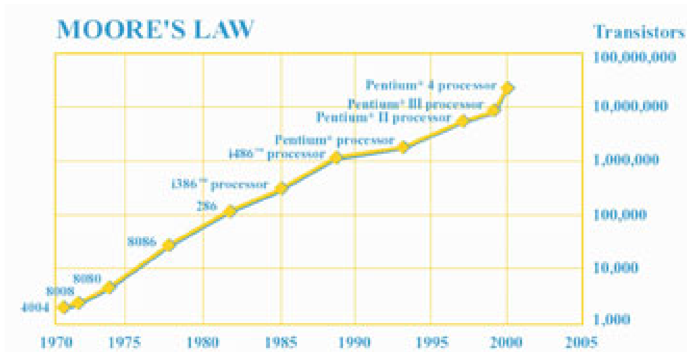


Figure 4.5: Evolutionary behavior is typically exponential

"grounding of semantics" can be left floating, as long as the "semantic atoms" are included in the data (see chapter 7). As compared to memetic systems, all relevant variables are visible in numeric form — they have the dimension of *money*. Assuming that the market reactions are consistent functions of the system state, including enough statistical features characterizing this state, self-contained balances can be defined. If the market has had time to converge to a higher-level balance, the neocybernetic principles can be applied for capturing the system state as a whole.

In today's world, the best proof of a new theory is the amount of money that can be earned when using it. Thus, the stock market offers a nice test bench — let us study a scenario to perhaps be tested in practice. Neocybernetic guidelines make the abstract modeling problem concrete and compact. First, to construct a model, statistical analysis that is based on the observation data only is sufficient; no complicated rule structures, etc., are needed to capture the balances. As discussed in 7.1.2, to capture the "cybernetic semantics", one also has to include the trends or derivatives of the variables among data in addition to the variables themselves. There are also guidelines for carrying out the preprocessing of the data: As explained in 4.3.3, the variables are scaled by their mean values, and only thereafter the mean is eliminated. This way, the cybernetically efficient variation is appropriately weighted; because the variables are always positive, such scaling is possible.

The neocybernetic model structure is then based on extraction of statistical dependencies among data in terms of sparse features (as explained in chapter 6). The algorithm assumedly reveals the market state in the framework of the neocybernetic market structure, showing the internal tensions within the market, making it possible to carry out predictions of the plausible developments.

4.2.4 Boosted evolution

As the properties of constructivistic systems are difficult to capture, changes in them are still more difficult to model. However, when constructing models for complex systems, and when trying to predict the future, such evolution processes are perhaps the most fundamental processes of all.

In the allocybernetic systems, individual humans implement the adaptation. Constructivistic systems can be designed and optimized explicitly, and when the agents are such intelligent, it should be easy to see where to go? However, as studied in the previous chapter (also see Section 4.3), the environment is unknown, and it changes as the system changes. The models need to be based on observations rather than on theories, and the adaptation process necessarily becomes iterative. There has to be enough time to observe the changing behaviors in the changing world. However complicated the environment is, it needs to be in balance with the system, and there exist some intuitions that are available here.

Humans are the agents that determine the variables and implement the enhancements — and the developments are caused by individual geniuses, making the evolution a very stochastic process. However, as seen from distance, details vanish: To penetrate the whole population, to become a truly revolutionizing change in thinking patterns in the large, any innovation needs to be accompanied by a large number of related breakthroughs in separate minds.

The memetic systems, too, seem to have their own internal dynamics. In allocybernetic systems, the agents do not experience physical hunger or other acute motivations, and different kinds of driving forces are needed to look for new frontiers. This mental imperative can be interpreted as "engineering spirit", curiosity that is boosted by greediness, resulting in objectives like *citius – altius – fortius.* These human aspirations, as seen from outside, become manifested as the systems "trying" to become somehow better: Faster, cheaper, more accurate, etc. Typically, the system goal is hypothetical, never reached — zero cost, zero delay, etc. — so that in this respect, the final balance is never reached, systems evolving forever. Momentarily, the cybernetic balance is determined by technical / economical / social possibilities and constraints.

In all its complexity, evolutionary processes can be abstracted in terms of the coupling coefficients q_i. Stiffness in the systems grows, coupling becoming stronger, q_i growing towards infinity. Why this happens — according to (3.13), the emergy transfer between u and $\bar{x}$ assumedly becomes boosted then, impedances getting lower, but another point of view is studied in chapter 10.

Very different phenomena affect the adaptation of the coupling coefficients, and this adaptation becomes a very random process. Parameters q_i are determined in other phenospheres, and there are many underlying variables and processes contributing, the cumulative outlook of behaviors becoming more or less continuous. In technical systems, when facing "designed evolution", developments are based on explicit investment calculations, economical pressures implementing balancing tensions, and smoothness and consistency in developments become explicitly underlined. Rather than studying q_i, it is easier to concentrate on $1/q_i$, typically having the unit of "price", "size", or "slowness" of a device. The final balance would be in zero, and as the process towards the balance can again be assumed to be a "next-level" generalized diffusion process, there will be exponential decay (see Fig. 4.5 for a manifestation of this "Moore's law"). Formally, this decay can be modeled as

$$\frac{d\ (1/q_i)}{\gamma' dT} = \frac{1}{q_i}, \tag{4.1}$$

where T yet slower time scale beyond t. No matter what is the prevailing level of q_i, the subsequent enhancements are relative to that level.

4.2.5 Hegelian megatrends

When trying to characterize extremely large systems, and when connection to concrete data is lost, the discussions necessarily become vague. Here, it is best to trust intuitions of established visionaries. The best explication of cybernetic ideas since Heraclitus (and equally obscure!) is given by Georg Wilhelm Friedrich Hegel (1770–1831). A more readable presentation of the "passions" of memetic spirits is given, for example, in [77].

Hegel was very influential in his time; it is essentially his ideas that are reflected, for example, in the writings of Johan Wilhelm Snellman, the inspirer of the Finnish national spirit. According to Hegel, history of mankind in general, and that of individual societies in special, is an evolutionary process[1]. In a way, Hegel can be seen as one of the first system theoreticians: Only the whole is consistent and a real thing, all partial explanations being illusory and deficient. Many of his thoughts can be interpreted in terms of cybernetic concepts — essentially, Hegel is speaking of *very complex agent-based emergent systems.* In the Heraclitus spirit, the essence is not *being* but *becoming.*

Specially, in a constructivistic system composed of human ideas, thoughts become diluted in the whole; true and false become intertwined, together constituting a consistent whole. The concept of "true" here contains logic and ethic considerations. What is more, everything is in change: The system becomes more and more complete in both logical and ethical sense. The results of human endeavors, or nation-states, are manifestations of the history, being — again applying modern terminology — relevant attractors of dynamical processes. Also Hegel's definition of what reality is like is very modern, emphasizing relevance: What is reasonable is real, and what is real must be reasonable.

Hegel emphasizes systems over individuals. For example, for him *freedom* is a contradictory concept: For individuals this is only freedom to follow laws (or some categorical imperatives), to make it possible for the larger system to become stronger and to develop further. The nation-state is not for its citizens, the citizens are for the state; it is a "person" of its own, deserving its existence over individuals.

The key concept in Hegelian philosophy is *dialectics,* or the idea of *theses* and *antitheses* (later employed by Thomas Kuhn). It is one dominating thesis in the society that finally finds an opposing antithesis, and together they form a synthesis. Contradictions do not collapse the Hegelian system; such seemingly illogical assumptions were attacked against by logicians. However, in the cybernetic setting, this all is quite consistent: The opposing theses determine tensions that together determine the dynamic balance that is necessary for the higher-level categories to emerge. The idea of dialectics was further elaborated on by Karl Marx and contemporaries; this is an example of how it is dangerous to apply rational reasoning without empiristic support — extrapolations easily result in irrational conclusions.

According to Hegel, all that fundamentally exists is really *mind* or some kind of *absolute idea.* The absolute idea is a consistent thought that "thinks of itself" — this is a poetic way of expressing a model being in statistical balance with

[1] Strangely enough Hegel did not yet foresee Darwin's work, never extending his studies to the biological realms

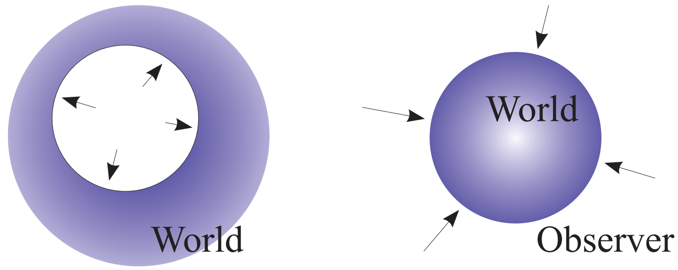


Figure 4.6: The traditional view of looking at complexity, on the left, and the cybernetic view, on the right. The system implements a mirror (or a lens?) that makes it possible project the infinite unstructured complexity onto a compact set of variables

the environment (see next chapters). The essence of this "Geist" or Idee" is similarly obscure as the Heraclitus' Logos is:

> *Der Begriff der Idee, dem die Idee als solche der Gegenstand, dem das Objekt sie ist.*

4.3 Quantification of phenomena

In the earlier chapters it was concentrations and other quantities that were easily coded in real-valued numbers. In complex environments, however, the variables generally cannot be quantified in such a straightforward way, and analyses have to be left on a more or less heuristic level. This applies specially to memetic systems — but also in economical systems, for example, even though it is money that makes values compatible, problems emerge if phenomena cannot be "moneyfied". Today's economic tradition seems to ignore everything that cannot be measured — but one should not deliberately limit one's analyses to approaches that already have been seen to be deficient when describing the complexity of the real world — otherwise, only a hollow formal system remains. Cybernetic considerations seem to offer new points of view here.

4.3.1 Mirrors of environments

As compared to traditional physical quantities, cybernetic variables are diffuse: They cannot be detected and quantified in an explicit way from "outside". They are fragile: Just as other emergent phenomena, any formalization of them misses their actual essence. They do not exist as independent entities, they cannot be isolated from their environments, as they are only relevant in interaction. In the cybernetic spirit, one can say that they are defined in terms of balances — employing the mechanical analogue, they can be defined in terms of the ratio between a "force" and the resulting "deformation".

To appropriately maintain the balance determining the observation, it is necessary to have the corresponding system connected in the environment. This means that the system disturbing the environment constitutes a "probe" that changes the potential tensions into actual observables. The infinite-dimensional complexity is projected onto a distinct set of variables. The results reflect not only the surrounding world but also the agents and their ways of seeing the world. Extending the Protagoras' statement, it is not only so that "man is the measure of all things", but it is *all cybernetic systems that constitute measures of their environments.* Remember the "Barnum effect": When there are enough variables, a consistent model can be constructed from practically any starting points (compare to the popularity of horoscopes, numerology, etc.).

This close coupling of the system and the environment means that the world also changes: Again employing the steel plate metaphor, affecting the deformability in one location affects the whole plate. When the system is completely cybernetic, there is constant stiffness in the observation points, so that the variation is pushed onto a constant level in *baru* — but simultaneously the system casts the variation onto another set of variables $\bar{x}$. In this sense, the environmental variation is mirrored onto the system (see Fig. 4.6). There can exist excess variation in the environment, but it remains hidden if there are no measurements. The world as it is seen is maximally supported, or spanned by observations; world is realized only after the measurement is carried out (compare to the "Schrödinger's cat"). Indeed, it is not only in the world of the simplest elementary particles where the measurement disturbs the system being studied — also in the other end of the continuum, when modeling extremely large systems, measurements alter the system being studied (or, more accurately, the measurement system alters the environment). In this sense, analysis of cybernetic variables is related to discussions concerning the general problems of *observer effect.*

Implementation of systems constitute the concrete "anchors" fixing the environment. Following Archimedes, one could say that "give me where to stand, and I will move the earth" — fix one point and the rest of the world will change to fit this constraint. As the world changes, the visible optimum state is dependent of the earlier decisions. The variability of evolutionary adaptation becomes easier to understand in this perspective: Because of the changing world, the optimum state is not predetermined after all (compare to Fig. 4.2). Whatever are the past developments, there are no dead ends, and further developments are based on the prevailing view of the world.

The extended capacities in intelligent agents make it possible to employ new variables, applying new interpretations. Variables in the memetic system are, for example, new concepts that change the ways how the world is structured. To find appropriate variables, intuitive understanding of the structure of the domain field is needed. The domain area expert typically recognizes the imbalance if there exists some, and innovations are introduced to compensate the tensions. In science, new theories are proposed — in economy, new products are proposed. A concrete example is needed here.

4.3.2 Cases of supply vs. demand

As an example of the "steel plate" analogue, study the *product market.* The market here is seen as the "product universe", unstructured entity, where the products constitute the contact points to customer needs, defining the set of

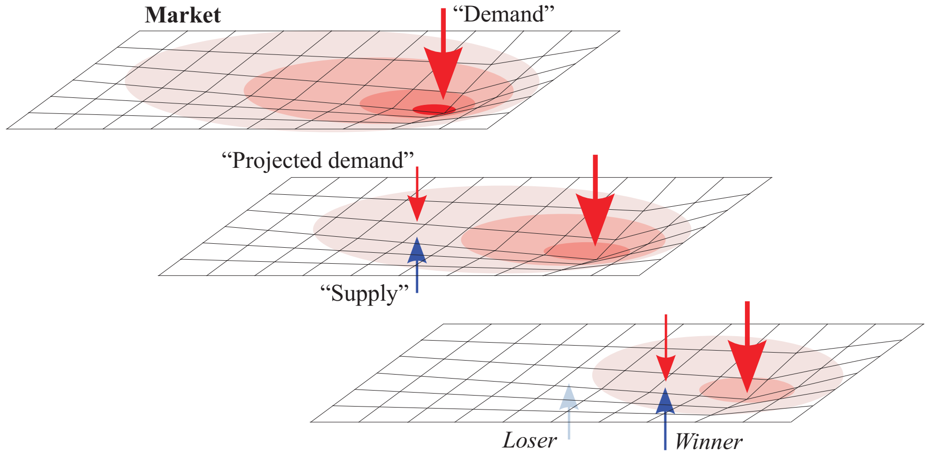


Figure 4.7: Exploiting the steel plate analogy (see Sec. 3.3.2). The unsatisfied demands (unobservable) are like forces that deform the balance market (here it is assumed that the force remains constant, so that the variance of this u_j is simply the square u_j^2). When a product is introduced (middle), it compensates some of the deformation, so that in that location the residue deformation after adaptation assumedly is only $1/\sqrt{q_i}$, as shown in (3.36)). When a competing product that in all respects better matches the demand is brought to market (bottom), the earlier product, masked by the new one, soon fades away: The deformations in that location remain below the level $1/\sqrt{q_i}$, and this product becomes "sub-cybernetic" (this behavior is verified also by simulations). Increasing the value of local q_i (lowering the price), or otherwise modifying the market (advertising the product, for example) can still help

quantifiable cybernetic variables. When this universe is presented as a concrete, visualizable entity, perhaps the simplification is not too radical (see Fig. 4.7).

The theories concerning the relationships between *supply* and *demand* are cornerstones of modern microeconomics [61]. There, the intuitions concerning elastic systems are clearly appropriate: For example, concepts like *price elasticity* are employed there. However, even though "demand" is a practical abstraction, it is difficult to quantify in general terms. It has been claimed that there cannot exist demand before there is supply — indeed, it is *Say's law* that puts it even stronger: "Act of producing aggregate output generates a sufficient amount of aggregate income to purchase all of the output produced". Now, the "forces" acting on the market are the potential demands, and the "deformations" are the actualized demands. It needs to be noted that, again, only balances are concentrated on, and it is assumed that in balance supply equals demand. In this sense, the vision of cybernetic economy is an idealized one.

In Fig. 4.7, it is shown how the abstract demand deforms the market. This demand is compensated by supply: A product is introduced — the properties of the product determine where it is located in the market. Some of the demand is projected onto this location, becoming a really measurable quantity. In the market domain, it is the products that are the probes quantifying the complex reality: The environment does "not exist" before there are the probes, so that the entrepreneurs restructure the world. In balance, it is the supply of the product that must equal its demand. If a more appropriate product is introduced, better matching the demand, the older product loses its support. It is the "stiffness" of the market that determines the properties of the equilibrium, and if a product is too "loose" as compared to the market stiffness, it is to finally vanish. Longer-range reformations of the steel plate as local changes take place can be interpreted so that the new products either complement or substitute other goods, changing their demand. If the demand is fixed, there can exist a monoculture after adaptation, but if there is variation — various "demand vectors" in different locations stochastically varying — there will be diversity in balance.

New products typically increase market stiffness, compensating still new locations of deformation, making the steel plate less compliant; this increase in stiffness is the natural route towards local evolutionary optimum. It is the products that implement this increase in stiffness, finally being distributed to compensate the external demands.

What is the physical interpretation of this stiffness, then? There are many factors that affect this coupling, but perhaps the most characteristic is *price,* or, actually, its inverse, so that the local q_i is proportional to the inverse of the price of the corresponding product. It needs to be recognized that for any selection of price there exists a balance — but the market is deformed, demands being redistributed accordingly. The higher the price is, the lower q becomes, and at some value this product drops out from into the subcybernetic region, finally fading away. If supply is well aligned with demand, and if there are no competing products, higher prices are tolerated.

The trivial goal that is always fulfilled for a cybernetic market is the local, product-wise balance of supply and demand, but the evolutionary goals on the local and global levels are different: Whereas an individual product provider tries to maximize the profit, maximizing the unit price or $1/q_i$, at the system level it turns out that the market becomes stiffer, q_i being maximized and price *minimized,* so that demand is better compensated by supply. And in an environment of independent distributed providers, if there is no monopoly, it is this system-level criterion that outweighs the local criteria, the individuals having to adjust themselves. How long the global-level evolution has proceeded is dependent of how mature the market section is, and how thoroughly the demand has been penetrated into.

Finding the "edge of the market surface" is of extreme importance for an individual product provider. This edge between the cybernetic and sub-cybernetic region can (in principle) be identified by experiments: Increase the price until (in balance) the net income does no more increase. It is the "effective actors" that determine the market structure, and if price changes do not cause changes in demand, the product is disconnected from the market surface. Here it is assumed that the product provider can respond to the whole balance demand; if there exist some hard limitations in production, for example, elasticity in this part of the market is lost. In the ideal case, the working economy becomes an image of the abstract market: Products are appropriately located, and they efficiently reflect the demands.

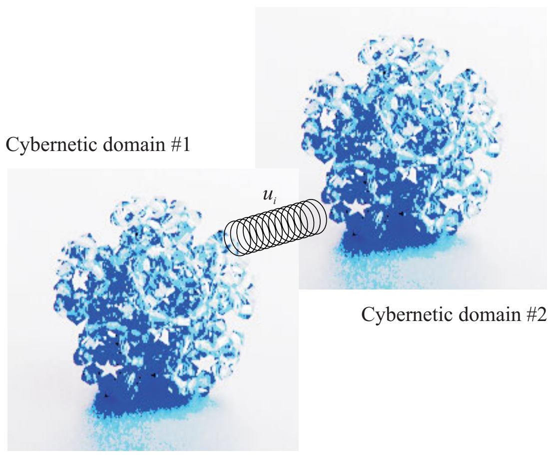


Figure 4.8: Abstract domains constituting cybernetic systems of their own are coupled by cybernetic variables

Where do the abstract demands originate from? Normally, such issues are not studied very much — this domain of the obscure human decision-making has traditionally been seen as something unmodelable. However, in the cybernetic framework, where abstract notions become quantifiable, it may be possible to go further. The human decision-making process assumedly is another cybernetic system where different kinds of tensions — desires — coexist, and this mental world can similarly be mapped in terms of a distinct number of cybernetic variables. The connection between the two cybernetic systems is instantiated by those cybernetic variables u_i that are common to both domains. And, in balance, it has to be so that the impedances in terms of q_i match in both systems: It is $1/q_i$ that determines the variation of the variable, and this variation is the same in both systems (see Fig. 4.8).

The above discussions can be generalized: One could say that *complex systems become observer-oriented systems,* the roles of “subject” and “object” becoming blurred. Indeed, they share the paradoxical properties of quantum mechanical systems, being like “Sch˚odinger’s cats”. In quantum mechanics, measurement makes the wave function collapse, revealing the single outcome out from the “cloud” of probabilities through the process of renormalization. Essentially the same intuition applies to the macroscopic as it applies to the microscopic: As studied above, the potential becomes actual through the process of becoming measured. In the extremely small scale, one disturbs the system being studied by accident, but in the extremely large scale, one has to disturb the system to make the measurements representative. The observer can only acquire relevant information about the system through becoming part of the system itself.

4.3.3 Towards different views of data

Two different, mutually incompatible approaches to seeing data have been studied this far: In this chapter, when studying populations, the models were *strictly additive* and globally linear, whereas in chapter 1, the connections among variables were *multiplicative* (models becoming locally linear after taking logarithms). Is there any possibility of again reaching homogeneity, so that it would be just one cybernetic model structure?

There are two levels of studying cybernetic systems: First, there is the *agent level,* being characterized by the individuals, and then there is the *population level,* where the effects of individuals cumulate. It is the individuals that are the actual actors, whereas behaviors on the population level are emergent. The appropriate way of seeing the environment changes when going from individuals to populations. The large number of individuals do not see the big picture, and it is not the actual level of variables that is of relevance. When trying to see the world from the viewpoint of an individual agent, it is first reasonable to divide the variable values (total activity of all agents) by the (average) number of the individuals — or, as the variable value is assumedly relative to the number of individuals, the variable should be scaled by its nominal value. As presented in the beginning of this chapter, the input variables, or the environment, can similarly be assumed to be consisted of other agent's activities, the same kind of prescaling should be extended to the input variables u, too.

But when looking the world from the point of view of an individual agent, additional modifications to data should be applied. As presented in the next chapter, it is *information* that plays the central role in cybernetic systems. This information is related to *variation* that the system experiences. Again, there is a big difference between the population and the individual agents: From the point of view of the whole long-living population, it is reasonable to study global long-term variations, that is, the behaviors of $\bar{x}$, whereas when studying the short-lived individuals that only can see a fraction of the time range of the whole population, the relevant variations are coded in the signals δx, transients around the local nominal values $\bar{x}$. This way, the whole range of behaviors in the data u, fast and slow, are modeled separately, together covering all of the variation. For the signals there holds $x = \bar{x} + \delta x$ and $u = \bar{u} + \delta u$, and for the accumulation of information, or "system memory" there holds

$$\mathrm{E}\{xu^T\} = \mathrm{E}\{\bar{x}\bar{u}^T\} + \mathrm{E}\{\delta x \delta u^T\}, \tag{4.2}$$

if $\mathrm{E}\{\delta x\} = 0$ and $\mathrm{E}\{\delta u\} = 0$. It turns out that this agent perspective — normalization by the nominal value and subtraction of the nominal value — is exactly the variable preprocessing that was proposed in chapter 1. Further, to make the models of the two levels compatible, also in the case of long-term modeling of populations, or when modeling the nominal values $barx$ and $\bar{u}$, it is reasonable to apply the same scaling: *Variables are to be divided by their mean values.*

It needs to be noted that this kind of scaling is *not typical* in practical data preprocessing — for example, this approach collapses for signals that have zero mean. However, such scaling can be done at least if the variables are strictly positive — as is the case with typical cybernetic variables.

When applying computational techniques for modeling data, determining the scaling of measurements is normally difficult, and some rules of thumb are normally applied: For example, the data is mean-centered and normalized to constant variance. The problem here is that when the modeling is based on (co)variations — as is the case when implementing PCA-like models, within or outside a cybernetic system — the scaling of variables is of huge importance, and formal variance equalization, even though being mathematically efficient, is not necessarily physically motivated. When the data dimension is high, these difficulties only become more acute. Now it seems that there is a more practical way of determining the scaling for cybernetic systems.

It seems that the behaviors and adaptations in cybernetic systems are dictated by the observations extracted from the environment. The key functionality in biological data processing seems to be redundancy elimination. When one now (assumedly) knows the principles of biological data processing, why could these data processing principles not be emulated, implemented outside the biological system, using computational approaches, to have a "cybernetic view" of the environment? When measuring phenomena, one uses the SI units, or some other technical standards — but such units do not necessarily have anything to do with the "natural" units, how the natural system sees the data. The measurements should become compatible if one applies special normalization for data: The measured values are to be divided by the nominal values. But one can extend these analyses of how data is seen.

Following the discussions in chapter 3, it seems that automatic data normalization is carried out for signals in cybernetic systems. There exist two extreme cases:

1. One could speak of "pre-cybernetic" data, if the measurements u are acquired from an environment where the feedbacks not yet have essentially modified the environment. Then it is the scaling of the data in the form $u = Mu_{\text{orig}}$, where $M = \mathrm{E}\{\bar{u}\bar{u}^T\}^{1/2}$, or $M = \mathrm{E}\{\Delta u \Delta u^T\}^{1/2}$, that is appropriate.

2. Then, one could speak of "post-cybernetic" data, if the environment $\bar{u}$ has been changed because of feedbacks in the way defined in chapter 3. Then (if q_i are assumed identical for different i) it is the scaling of the data in the form $M = \mathrm{E}\{\bar{u}\bar{u}^T\}^{-1/2}$, or $M = \mathrm{E}\{\Delta u \Delta u^T\}^{-1/2}$, that is appropriate to reach the "cybernetic" view of data.

In the former case, it is essentially the framework of principal component analysis (PCA) or principal subspace analysis (PSA) modeling of data that applies. However, in the latter case, it seems that *all variation-based information is ripped off the data,* and PCA-based methods collapse if only such "cybernetized" data is available. Eliminating the covariance structure means *whitening* of the data, and there is a connection to *independent component analysis (ICA)* or *independent subspace analysis (ISA)* here. In a post-cybernetic environment, where covariance structure has been ripped off it is the higher-order statistical properties only that remain available in the data; these properties can be made visible for PCA-type algorithms when special kind of nonlinearity is included in the structures (see *fourth-order blind identification (FOBI)* in [41]). Assumedly there is a continuum between the extremes, and the above analyses are clearly unsatisfactory.

As it turns out after closer inspection in chapter 6, where different views of seeing data are further elaborated on, rather than pure PCA or ICA, it is *sparse coding* that is being implemented by the cybernetic system. This coding is more robust against scaling, etc, and, what is more, such more complex codings can be reached without introducing extra nonlinearities.

Now, assume that all phenomena have been successfully quantified. It was *information* that turned out to be crucial for appropriately quantifying behaviors in cybernetic systems, and as it turns out, this concept is useful when abstracting away the details of the domain field. In the next chapter, the cybernetic domains are seen from yet another point of view, employing new and powerful concepts.

Level 5

Role of Information in *Model-Based Control*

The neocybernetic analyses started from simple, reductionistic studies. As the analyses were extended to wider-scale systems, the focus points changed, and new points of view were employed. However, to reach the truly holistic view, yet other interpretations are needed. No new concepts are needed — it turns out that one only has to exploit familiar concepts in new ways. For example, the term "information" has been used routinely, but only intuitively: This is one of the key concepts that open a completely new perspective towards cybernetic worlds.

Many of the cybernetic intuitions become explicitly quantifiable in the neocybernetic perspective. It turns out that when the powerful tools of *control theory* become available, a beautiful new world becomes visible.

5.1 Another view at emergy

The concept of *emergy* was presented in chapter 3, and it turned out that the evolutionary processes could be formulated in that framework. Emergy, the effect that is interpreted as tension, essentially differs from the concepts of energy or power: It is *deviation from the expected* that is crucial — or *information.*

5.1.1 Information vs. noise

Ross Ashby coined the *Law of Requisite Variety* in 1952:

> The amount of appropriate selection that can be performed is limited by the amount of information available.

This is a deep observation — but very "Heraclitus-style", being left obscure. The concept of information is left vague here, and the consequences remain unclear. However, speaking of information seems to offer just the appropriate connotations. To make it possible to efficiently apply mathematical tools for analysis of information flows, the basic concepts necessarily have to be defined in an accurate manner. So, information in the environment is presented by the data, and this data is coded in real-valued signal vectors. How is information manifested?

One is facing a *reverse engineering problem* here: It is known what the cybernetic system (assumedly) does with the data if acquires, and when employing the new terminology, it is assumed that *information is what information processing in natural systems does.* One has to hope that the intuitive notion of information matches with what a cybernetic system is accomplishing. In chapter 3, it turned out that the weighting matrix in the pattern matching is

$$W = \mathrm{E}\{\Delta u \Delta u^T\}. \tag{5.1}$$

This means that data is weighted by the correlation matrix when evaluating matches among patterns: The neocybernetic system must see *information in variation.* The corresponding models are fundamentally based on correlation matrices — principal subspace analysis is just a way of formally rewriting and redistributing this correlation information. The correlation matrices contain atoms of information, entries $\mathrm{E}\{\bar{x}_i \bar{u}_j\}$ revealing cumulated pairwise (co)variations among variables, or *mutual information.*

The correlations and covariances have traditionally been exploited in modeling — what is new in neocybernetic models? Covariances and variances are simple measures for information, being easily expressed and exploited, and they are the basis of modern identification and minimum-variance approaches in systems engineering. The key observation when comparing cybernetic data processing to traditional identification was studied already in chapter 2: Traditionally, when doing parameter fitting applying maximum likelihood criteria for Gaussian data, the approach is opposite — variation is interpreted as something to be avoided — and the weighting matrix is the *inverse* of (5.1). Variation is interpreted as *disinformation,* or noise.

As Gregory Bateson more or less intuitively puts it [7]: "Information consists of differences that make a difference". It is not whatever variation that is thought to be interesting in cybernetic systems: It is *covariation* among data items that is not sensitive to surface-level phenomena like measurement errors, but reveals the underlying common sources or deep patterns. No matter what is the application domain, this covariation is always assumed to be interesting. The role of the cybernetic machinery is to capture the information in compressed form with minimum number of parameters; the correlation matrices that are constructed are essentially storages of the mutual information among data. When the basics are simple and efficiently implementable, accumulation of the information structures makes emergence possible (see chapters 7 and 9).

Such a mechanistic view of information is, however, somehow incomplete. The concept of information also carries something veiled and mysterious that is related to knowledge and *meaning.* One should not lose the power of intuitions; indeed, the concept of information gives tools to attack the problem of *relevance,* too.

When applying Shannons information theory (or Kolmogorov / Chaitin (algorithmic) information theory), the definition of information is strictly syntactical. There is no domain area semantics involved, and thus extreme universality is reached. However, some paradoxes remain: What you expect, contains no information, and it is noise that has the highest information content. When applying the neocybernetic view of information, semantics (in a narrow, formalized sense) is included in manipulations, making the analyses non-universal — but there is *universality among all cybernetic systems.* The approach is intuitively appealing: What is expected, is the most characteristic to the system, and uncorrelated noise has no relevance whatsoever. Capturing the cybernetic semantics and modeling of knowledge is studied in more detail in chapter 7.

5.1.2 State estimation and control

A cybernetic system is a "mirror" of its environment, optimally capturing the information there is available. This is not merely a metaphor — note that the formulas in chapter 3 can be given very concrete interpretations:

- **Model.** It turns out that the neocybernetic strategy constructs the *best possible* (in the quadratic sense) description of the environment by capturing the information (covariation) in the environmental data in the mathematically optimal principal subspace based latent variables:

$$\bar{x} = \left(\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}\right)^{-1} \mathrm{E}\left\{\bar{x}\Delta u^T\right\} \; \Delta u. \tag{5.2}$$

- **Estimate.** It turns out that the neocybernetic strategy constructs the *best possible* (in the quadratic sense) estimate of the environment state by mapping the lower-dimensional latent variable vector back onto the environment applying the mathematically optimal least-squares regression formula (2.22):

$$\hat{u} = \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T \left(\mathrm{E}\{\bar{x}\bar{x}^T\}\right)^{-1} \; \bar{x}. \tag{5.3}$$

- **Control.** It turns out that the neocybernetic strategy integrates modeling and estimation to maximally eliminate variation in the environment:

$$\tilde{u} = u - \hat{u} \tag{5.4}$$

Even though the operations are represented here in such compact and centralized form, all operations are strictly local, and the represented net effects are only visible as emergent phenomena; for example, the feedback part is implicit. Implicit feedback makes the mappings more conservative: For example, the estimate between $\bar{x}$ and u is indeed implemented applying the regularized least squares formula (2.20), with the role of the regularization parameter q now inverted. The issue of modeling Δu rather than u directly is studied in Sec. 5.2.1; when q increases, u and Δu approach each other what comes to the n most significant eigenvalues.

The above observations mean that a cybernetic system implements *model-based control* of its environment. In terms of information as defined above, this control

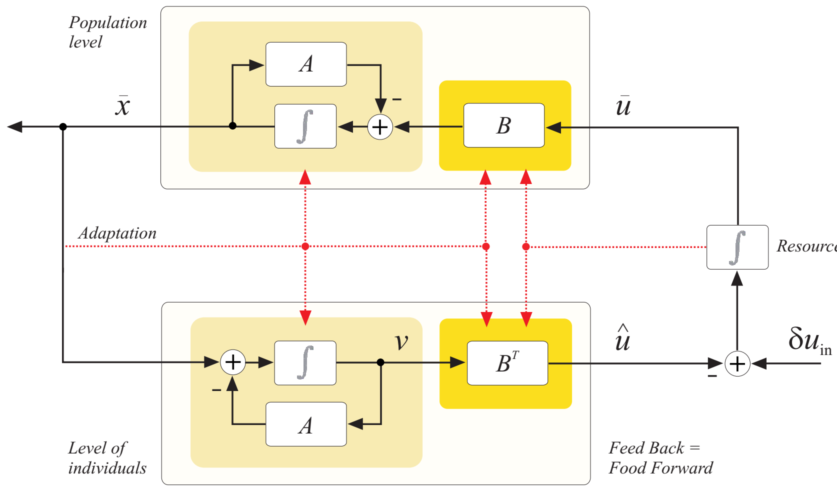


Figure 5.1: Cybernetic system seen through the eyes of a control engineer

is the *best possible.* However, note that the controller is defined as a static structure, control emphasis being shifted from dynamic transients to stationary statistics; the hypothesis here is that however the information acquisition is implemented (for example, as a time-series structure resulting in traditional dynamic control structures; see chapter 7), the cybernetic system maximally compensates that information. The implemented control is far from trivial: It constitutes a multivariate controller where the n most significant variation directions are equalized (or nullified). The symmetric structure of the modeling / estimation loop reminds of Heraclitus' words: "The way up and the way down is the same" (see Fig. 5.1).

In the selected framework, age-old intuitions become concrete. Indeed, the control intuition — cybernetic systems do control — has been clear since Wiener, but the mechanisms have been unclear. Ross Ashby also coined the *Law of Regulatory Models:*

> Regulator must not only have adequate amounts of variety available, but also be or have a *homomorphic representation* of that system.

Since that, the same idea has been known in the field of control engineering as the *internal model control* principle: A controller must contain an (inverse) model of the system to be controlled. Still it needs to be emphasized here that whereas traditional control is always centralized, based on some "master mind", now the control structures are completely distributed: The starting point was local level feedback controls, but the final result is global level feedback control.

Ross Ashby also states that "for appropriate regulation the variety in the regulator must be equal to or greater than the variety in the system" (Ashby's "regulator" being the system, and "system" being the environment). However, here his intuition is *wrong.* The capacity of the cybernetic system must be *less* than that of the environment. If there is no scarcity of resources in the system,

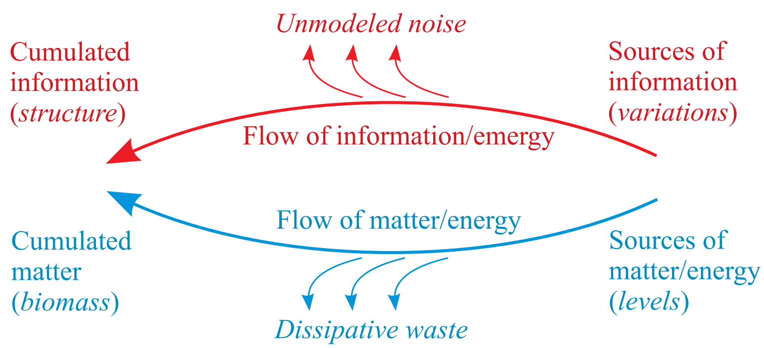


Figure 5.2: Abstract flows in a cybernetic system

no compression — or modeling — needs to take place, and no cybernetic system can emerge. It is the environment that dictates the terms.

5.1.3 Flows of information and matter

Information is also the common denominator capturing the essence in cybernetic systems. Everything that affects the behaviors can be seen as visible (measurable) variation or information; it is information that is being controlled in the environment, and information is being cumulated in the model. Further, information makes different models commeasurable, and information determines the semantics and goals of the system. Yet another viewpoint to the role of information is available here.

The feedback part in the closed-loop structure in Fig. 5.1 is only an abstraction: It does not correspond to a separate real process because it only represents the non-ideality of the information transfer. It is interesting to note that for the closed loop control structure to emerge, two different kinds of processes need to co-operate — first there is the information flow into the model, and then there is the material flow dictated by the model. Without the other flow the other could not exist either. One could say that a cybernetic system constitutes a *marriage mind and matter,* combining these two incompatible dualistic viewpoints (see Fig. 5.2).

In the figure, there are the two flows shown separately: On top, there is the flow of information (or emergy), and on bottom, there is the flow of matter (and energy). Most is wasted — in information flow, the uncorrelated noise becomes filtered, whereas in material flow, it is the dissipative losses that do not get through into the higher-level system. Note that it is often assumed that it is these dissipative material flows that are the manifestation of complex system dynamics [64] — now these are just a side effect. It is the information in the environment (or *variations* in the data) that dictates the structures within the higher-level system, whereas it is the matter (or actual *levels* in the data) that cumulate as some kind of biomass within this predestinated structure of some kind of populations. Whereas the traditional matter and energy oriented views emphasize the level of dissipation, levels of flows being the most essential, in the neocybernetic information oriented perspective constant flows are seen as trivial and not interesting from the point of view of emergent structures.

One could even say that the cybernetic model to some extent captures the Platonian *ideal* beyond the changing world.

5.1.4 Different views at the environment

Here, an example of what are the benefits of applying concrete definitions for concepts is presented. And, again, it is visualized how the fact that real systems are not ideal brings sophistication in the discussions; things do not necessarily become more complex, but new nuances are introduced in the models, and deeper understanding can be reached.

It is assumed that in a long run an *evolutionarily surviving system exploits all information it can see:* Being capable of efficiently exploiting the resources is a prerequisite of surviving in an environment, successful systems are the most active in acquiring for more and more information. This optimality assumption makes behaviors in an environment more or less unique and predictable. When modeling such systems, the optimization task is somewhat trivial, when constraints are given. The interesting challenge is to understand the different mechanisms for information acquisition; why there can still exist different kinds of systems in the same environment, can be studied by assuming that there are different kinds of constraints in the information capture process, and different systems see the environment in different ways. Here, a special aspect is concentrated on: *There can be differences in how systems remember their experiences.* Within the introduced framework these issues have a compact "vocabulary" (distribution of information is further elaborated on in chapter 6).

This far, the expectation operator has been employed in a sloppy way: Indeed, expectation is a mathematical abstraction that cannot be measured, it can only be estimated using the measurement samples. Accurate determination of expectation would necessitate an *infinite* number of samples — this is clearly impossible at least in the changing environments. Instead of employing the mathematically accurate definition, define the "expectation estimate" be an (exponentially) weighted average over the past observations:

$$\frac{d\hat{\mathrm{E}}\{\bar{x}_\mathrm{s} u_\mathrm{s}^T\}}{dt} = -\gamma_\mathrm{s}\hat{\mathrm{E}}\{\bar{x}_\mathrm{s} u_\mathrm{s}^T\} + \gamma_\mathrm{s}\bar{x}_\mathrm{s} u_\mathrm{s}^T. \tag{5.5}$$

Now, there is an exponential "forgetting horizon" what comes to the covariance estimates: Newest observations are best remembered, whereas old experiences fade away with time. In the similar manner, assume that there is inertia and forgetting taking place in all data processing in the system, so that also the incoming data is seen through such filter:

$$\frac{du_\mathrm{s}}{dt} = -\mu_\mathrm{s} u_\mathrm{s} + \mu_\mathrm{s} u_\mathrm{in}, \tag{5.6}$$

Here, u_in is the original input supplied by the environment, and u_s is the filtered input actually seen by the system; the parameters $\mu_\mathrm{s} > 0$ and $\lambda_\mathrm{s} > 0$ are the filtering coefficients, higher values meaning fast forgetting. This extension makes it possible to take variation structure in time domain into account.

Such linear time domain filtering can most efficiently be represented and analyzed in *frequency domain.* It turns out that information can directly be analyzed in terms of *power spectra.*

To illustrate this, observe that for the Laplace-domain signals $\bar{X}$ and $\bar{U}$, one can express the filtering of signals as $\bar{X} = F\bar{U}$, where the *transfer function* for the first-order filter (5.6) as

$$F(s) = \frac{\mu}{s+\mu} U_{\text{in}}(s), \tag{5.7}$$

and, further, the power spectrum of this becomes

$$H(\omega) = \frac{\mu^2}{\omega^2+\mu^2} H_{\text{in}}(\omega). \tag{5.8}$$

This reveals that the transfer from input power (information) to the power that is actually experienced by the system is a function of angular frequency ω. For low frequencies, $H(\omega) = H_{\text{in}}(\omega)$, but beyond the cut-off frequency μ_{s}, the experienced power decays linearly when studied on the log/log scale.

The filtering effects are visualized in Fig. 5.3 — there it is shown how the information content of a signal can reside in different frequency regions. Frequencies above the cut-off frequency μ_{s} are seen as noise by the system, and gets ignored altogether. Frequencies below that are seen, but assuming that $\mu_{\text{s}} > \gamma_{\text{s}}$, they do not get cumulated in the system's structures — these frequencies are only filtered, or "manipulated" by the cybernetic system. Only variation in the darkest area in the figure becomes cumulated in the model (or in the covariance matrices). Too high frequencies are invisible altogether to the current system, leaving there room for other systems to flourish; but also in the lower frequency range ("environment"), there is competition; even though such signals are visible to the system, there exist probably more customized systems eliminating that variation. The net effect is that the system concentrates on band-limited signals only, signals in other frequency ranges being interpreted either as noise or as constant values — both containing zero information in the cybernetic perspective. The observation from chapter 4 (the behavior of the nominal state, and deviations around it can be modeled by separate systems) can thus be extended and made better quantifiable.

Such differentiation among systems, makes them mutually dependent. Specially, if the lower-range model changes — as it necessarily does in practice when time goes on and the slow phenomena become better visible — the higher-range systems need to adapt to this changing environment; and the needed adaptations can be rather abrupt. Discontinuous changes in the environment are magnified in the subsequent systems.

5.1.5 Cascades of trophic layers

Information is the "nourishment" for systems. It does not matter if the driving force is *loss* of some resource (as when allocating staff labor) or surplus: Posi-

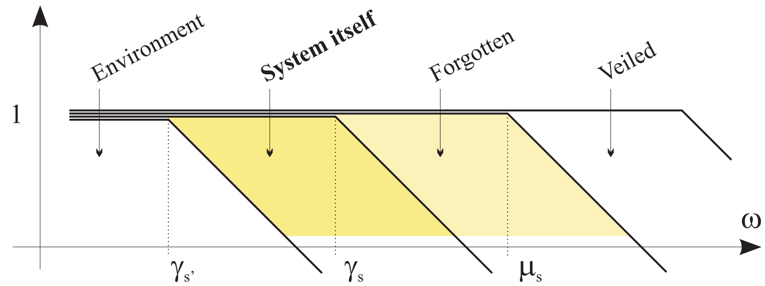


Figure 5.3: Different systems operate on different time scales

tive or negative, the effects are the same. A cybernetic system sees information (emergy) as resources available in the environment, and there is hunger for this information. Again, this sounds teleological — but if some system applies this strategy by accident, it immediately has evolutionary benefit in terms of increasing resources. There is no guiding hand needed — but it is like with Gaia: Even though all behaviors can be reduced to lower levels, simplest models are found if stronger emergent-level assumptions are applied. It turns out that this eternal hunger for information has resulted in very ingenious-looking solutions for reaching more and more information, and, to achieve the necessary sophistication, the systems have typically become ever more complicated. The issues of such information pursuit are studied more in chapter 6.

The systems are hungry, but they are not greedy. Whereas a system exhausts variation in its environment, there is the same variation inherited in the system itself (remember that PCA model maximally relays variation to its latent variables). This gives rise to a *cascade* of trophic layers: Another system can start exploiting the variation that is now visible in the system (being part of the environment as seen by the other systems). When the next trophic layer has been established, there is room for a yet higher trophic layer, etc.

In nature, the basis for all life is the Sun. However, the “non-informative” sunlight alone is *not enough* for cybernetic systems to make them flourish — or, indeed, it is not enough to make them emerge in the first place. Additionally, there are first the physical processes (planets orbiting and rotating) generating more or less cyclic variation in the physical variables, causing temperature gradients. These give rise to second-level chaotic processes: When there are temperature gradients, it is the highly nonlinear Navier-Stokes type equations that produce increasing amounts in randomness in the variables, as being manifested in climatological phenomena, etc. Now, the arena is free for cybernetic systems to start exploiting this non-trivial information; after the information already is there, linear processes are enough to utilize it. The input variables for the lowest-level cybernetic systems (plants) are temperatures, nutrients in the soil, rainfall, etc. On the level of herbivores, it is then the spectrum of plants to forage on, and after that are the carnivores foraging on each other. All loose information seems to give rise to new systems, and, in a way, this can be described as “panspermia”. As the number of species increases, the complexity also increases, as the subsystems become more and more interlinked: There emerge pests and diseases to exploit the variety, too. It is only natural that at some stage the lower level species adapt to utilize the higher-level biomass

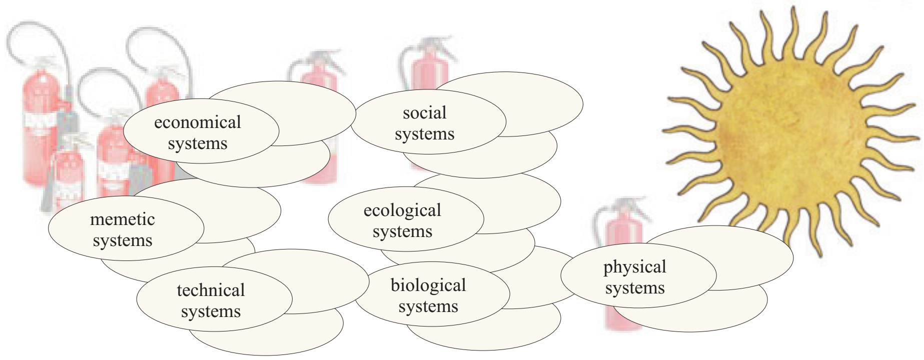


Figure 5.4: Systems in different phenospheres try to extinguish the fire (Heraclitus' *Logos*)

— without recirculation the (dead) biomass would cumulate indefinitely in the resource vector. This makes it a cycle, and finally the natural circulation is established as a consequence of locally controlled information exploitation.

When the succession of systems evolves, the highest-level systems can appear in very different phenospheres. Above the biological systems, there are all the man-made constructivistic systems — but they still live, after all, on the variety resources of the nature: For example, take the scientific systems. Without the simpler cybernetic systems there would be no natural sciences, and without more complex cybernetic systems, there would be no social sciences; without uneven distribution of nature's structures there would be no need for explanations. What science explores, technology exploits — environment being exhausted as a result of such loop. All systems finally try to exploit (or *eliminate* when seen from another point of view) the Sun's fire, either directly or indirectly[1]. Indeed, sun-worship is among the oldest rites. And Heraclitus said that the underlying principle in nature is *fire.* However, in the cybernetic perspective, this is not the key point: It rather seems that the goals of nature could best be explained in terms of a *fire extinguisher.* (see Fig. 5.4).

When the internal inertia in the cybernetic systems is taken into account, one can think of the information transfer between subsystems as some kind of a potential flow from trophic layer to another. There is a "structured leakage" in the information reservoirs; this can also be characterized as "directed diffusion". The subsystems are like (generalized) "ideal mixers" — mixing information (note that the flows are not scalar variables but vectors). As linear systems, the cybernetic mixers can be grouped in different ways; the subsystems seem to be tightly connected and they always define a network, however they are regrouped. When more and more layers are introduced, the ecosystem becomes more and more continuous and smooth from the perspective of information distribution – becoming a lumped parameter approximation of a parabolic partial differential equation (PDE) diffusion model. The evolutionary process of sophistication continues until there are incompletely exploited reservoirs of resources available.

[1] Or, actually, *primus motor* is the fire from the Big Bang: The geological conglomerations and variations in soil properties that also have to be seen as cybernetic resources are not caused by the Sun

Finally the "landscape" should become smooth with no sudden drops, no matter how the intermediate levels are constructed. Changes in resources get filtered when they spread among the systems.

When looking at the wealth of systems that exist to implement the extinction of fire, one cannot help thinking that *the right hand does not know what the left is doing.* It is not about an "intelligent designer"; one could speak of a "hardworking blunderer" instead[2]. The philosophical question is not where the diversity comes from, but why there is *something* instead of *nothing.*

5.2 Control intuitions

Even though truly complex systems cannot be easily quantified, they must share the basic principles: If a system is to remain consistent, there has to exist the balance of tensions deep inside. Qualitatively, identical intuitions apply. When the control notions are employed, it turns out that there are many intuitions directly available for analysis of the behaviors in cybernetic systems — and *vice versa.*

5.2.1 Rise and fall of adaptive control

Adaptation is the key property in truly cybernetic systems, meaning that they are *adaptive control systems,* trying to implement more efficient controls based on simultaneous observations of their environments [3]. If one has control engineering background, one can immediately understand what happens in a truly cybernetic system then: Adaptive controllers are notorious in control engineering, as they can behave in pathological ways. The reason for the "explosions" is *loss of excitation.* Good control eliminates variation in data — and after this there is no information where the model tuning can be based on, and gradually the model becomes corrupted. After that, when the model is no more accurate, the variation cannot all be eliminated, and the control performance can be very poor. But as the control fails, the variation cannot any more be suppressed, and there will exist information in observations once again. The model starts getting better, and after that the control gets better, and the cycle of good and bad closed-loop behavior starts again. This kind of oscillatory behavior is typical in loops of simultaneous model identification and model-based control. This result is paradoxical: Pursuing good balance on the lower level results in high-level instability.

Is it reasonable to compare complex cybernetic systems to simple controllers? This question is motivated as the processes in real life systems are so much more delicate — but still there is some resemblance in the emergent behaviors. Compare to ancient empires: It seems to be so that there is a life-span for all cultures, after which even the strongest civilization collapses. Why is that? For example, during "Pax Romana", there were no enemies, and the once famous Roman army became ruined, morally and otherwise – and then there was a

[2]It would take a truly "intelligent" agent to streamline the natural systems. God forbid that there should be such re-design efforts ...

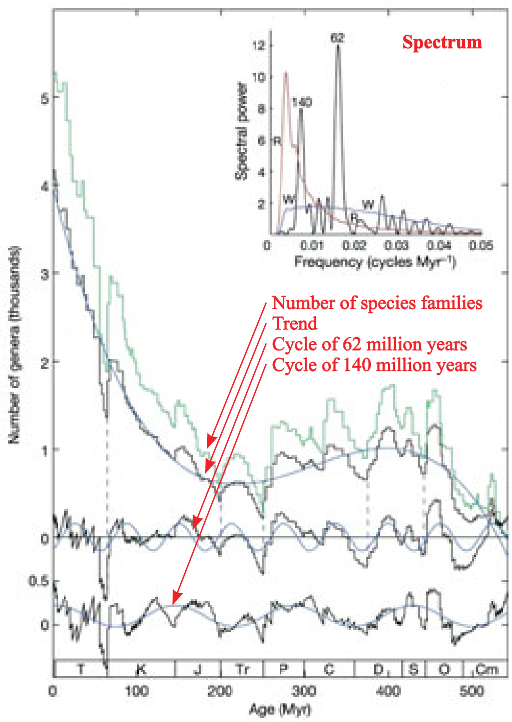


Figure 5.5: For some reason, the ecosystem has periodically become turbulent (diagram adopted from [65]). Note that dinosaurs became extinct about 62 million years ago

collapse after severe disturbances[3]. And this increase of sensitivity does not only apply to human societies (see Fig. 5.5): For some reason, massive extinctions seem to take place in 62 million year cycles [65]. Do you need some meteors to explain extinctions — or is this simply because of evolution dynamics? It seems that current explanations to collapses in general prefer simple solutions (see [22]).

Extreme optimization in some respect results in worsened fitness in changing conditions, and a collapse of the highly specialized subsystem (or the whole ecosystem) is possible. Of course, nature has developed mechanisms to cope with this challenge. For example, in natural systems, there are multiple local minima simultaneously represented. Different species are optimized with respect to their local view of the environment, and as such a pool of structural alternatives is maintained, not the whole system needs to collapse when the environment changes as suitable candidates also exist.

To reach smoother behaviors, there exist other alternatives in addition to the *multiple model approach,* and, again, the technological experience can be exploited here. In control engineering, techniques have been developed to tackle with the adaptive systems: One of the basic techniques is to *add noise* to introduce fresh information in the closed-loop system, preventing the control from becoming too good. A more sophisticated technique can be seen as an extension of this: The controls are designed to artificially make the system roam through

[3]But explicit emphasis on the army results in the Soviet-type collapse: If there is no real need at some time, such investments are cybernetically non-optimal, meaning that the system cannot outperform its competitors in other fields in the evolutionary struggle

the admissible region, thus exciting the modes, and mapping the responses and latent dynamics. For example, in complex industrial plants such control strategies are commonplace, reagents being added until some specific criteria are reached, and after that reagents being reduced until some other criteria are reached. Of course, this results in oscillation (limit cycles) in the closed-loop system, and thus in variability in product properties — but, regardless of its limitations, such cycles are employed also in real natural systems, caused by, for example, the *cell cycle* in cultivations. Formally, a well-behaving system is seemingly permanently on its stability limit.

It seems to be always so that the optimality goal has to be relaxed to reach *good* behavior. The above solutions — messing the control up with more or less stochastic or deterministic noise — add the element of randomness and unpredictability in the system as seen from outside. However, there seems to exist yet another elegant technique that is inherently applied by the natural cybernetic systems. The most important ingredient here is again trivial, caused by the nonideality of nature: It is the *stupidity of agents* that facilitates the emergence of sustainable systems.

5.2.2 Paradox of intelligence

As compared to traditional adaptive controllers, the cybernetic strategy where the feedback is implemented implicitly through the environment, results in "gentle" adaptive control, form of *buffering,* where the variation is not fully eliminated, and the closed loop behavior does not become pathological: There will always remain enough excitation in the signals. One could also speak of *passive control* as only *attenuation* of signals takes place; how near complete elimination of excitation one goes, is determined by the coupling factors q_i. This is because it is Δu rather than the estimate u itself that is being eliminated from the input data, making the overall system evolutionarily stable and sustainable. But such control, leaving some of the input uncompensated, is technically not optimal — and cybernetic systems always pursue better controls ...

Indeed, getting too ambitious, implementing extreme optimization, and full exploiting the information completely wiping out excitation, is also a possible scenario in a cybernetic system — if the system is sophisticated enough. This kind of invasive, fully compensating control can take place if the agents realizing the control are "too smart", implementing the feedbacks explicitly, actively, rather than waiting for the environmental reactions.

To implement such extreme optimization, the different signals have different roles as seen by the agents: The inputs and outputs need to be functionally separated from each other, meaning that the system necessarily has more sophisticated, predetermined structure, as seen from outside. When the competition among agents is explicitly taken into account, one can start the modeling from (3.4) and write

$$\frac{d\,x}{d\,t}(t) = -\Gamma A\,x(t) + \Gamma B\,u(t). \tag{5.9}$$

Here, the gradient expression is extended by taking into account that the diagonal Γ makes it possible for agents to have differing adaptation speeds. Now, when defining

$$A = \Gamma \mathrm{E}\{\bar{x}\bar{x}^T\}, \qquad \text{and} \qquad B = \Gamma \mathrm{E}\{\bar{x}u^T\}, \tag{5.10}$$

one changes the original feedback structure in chapter 3 only minimally. Essentially all signals are handled identically, and weight adaptation is identical for all signals — but there is a twist: If a signal is known to be recirculated, if it belongs to the x variables, its value is additionally multiplied by -1, as shown in (5.9). This is what it takes to actively implement the negative feedback: The agents only need to distinguish between "positive" and "negative" inputs, or information about resources and competitors, respectively. Implementation of the explicit feedback in this way results in combined Hebbian/anti-Hebbian learning (see [92]). The matrix A now defines the communication (or, at least information transfer) among the agents. In large systems, the size of this matrix (having n^2 elements for an n-agent system) can become considerable necessitating structured coordination of signal transfer. In any case, if u varies slowly, the steady state for x is defined through the mapping matrix

$$\phi^T = \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}u^T\} \tag{5.11}$$

so that $\bar{x} = \phi^T u$. From discussions in chapter 3, when Δu is now everywhere substituted with u, it is clear that the columns in ϕ span the principal subspace of u, and PSA is implemented explicitly for u. Remember that as the feedback in the "smart" structure is implicit, all signal manipulations taking place within the system, the input data is not disturbed. In this sense, the signal transfer is idealized, information theoretic, assuming that observation can be implemented without exhaustion of the signal source. Also in this sense, the smart agents assumedly operate on a higher abstraction level, not being bound to their immediate surroundings. The disadvantage is that as the input signal is not touched, no control is automatically implemented. In the model-based controller structure in Sec. 5.1.2 two items are also changed:

- The model becomes

$$\bar{x} = \left(\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}\right)^{-1} \mathrm{E}\left\{\bar{x}u^T\right\}\ u. \tag{5.12}$$

- The estimate becomes

$$\hat{u} = \mathrm{E}\left\{\bar{x}u^T\right\}^T \left(\mathrm{E}\{\bar{x}\bar{x}^T\}\right)^{-1}\ \bar{x}. \tag{5.13}$$

However, cybernetic systems are for control purposes — so, if the feedback structured are separately hardwired, applying the "smart" model for explicit control, all available variation in u is exhausted. This results in all the familiar problems of traditional adaptive control. When you can optimize, you typically do it, even though optimal is the enemy of good in the sense of robustness and sustainability: "It is hard to be humble when you are so strong"!

But there are also benefits when feedbacks are optimized — the system can truly be smart, and there is evolutionary advantage. Unnecessary competition can be avoided, resources can be allocated by negotiation (more or less democratically), and the agents can concentrate on more productive issues. As a consequence, a

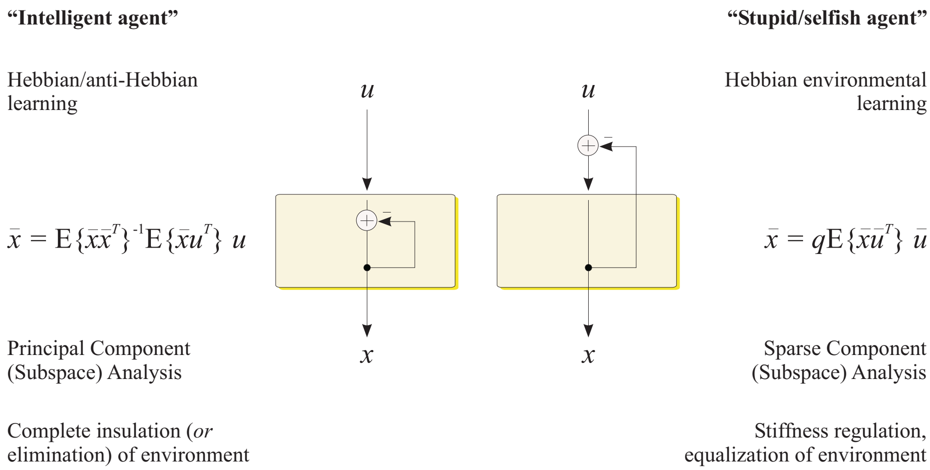


Figure 5.6: Two learning strategies, two ways to see the world and change it. In a system of "intelligent" agents, the interactions among the competing actors are taken explicitly into account, being integrated in the adaptation strategy, whereas in a system of "stupid" agents, adaptation takes place in the direction of visible resources, the interactions becoming evident only implicitly through the exhaustion of the environment (details of differences in input coding are presented in chapter 6)

welfare state need not necessarily be less efficient than a pure capitalist economy — assuming that the model of the (changing) environment (legislation, etc.) remains up-to-date. The two types of feedback implementation strategies are illustrated in Fig. 5.6.

5.2.3 Contribution in inverse direction?

It is not only so that control intuitions would be applicable in analysis of cybernetic systems — there is contribution in the inverse direction, too. It may be that the locally adapting controller schemes could make it possible to implement controls that cannot have been imagined this far. The applications can range from sensor fusion to agent controls and complex networks in general. What is more, the cybernetic systems of humans, the process operators, can perhaps be integrated in the cybernetic models of the processes — issues of "human factors" can perhaps be addressed fluently in the same modeling framework.

Today's main challenge in control engineering is understanding complex automation systems: How emergent properties like *robustness* could be seen from designs, how to find analysis and synthesis methods to address qualitative plant properties?

An industrial plant is "first-level cybernetic" because there are controls implemented so that it can sustain environmental disturbances and it (hopefully) finds a new balance if the conditions change; the industrial process can be seen as an “artificial cell” with its own metabolism, “eating” the raw materials and giving out the products. Applying intuitions concerning natural cells and their robustness, one would like to extend from the first-level to second-level of cybernetics, so that higher-order statistical balance between the system and its environment would be reached, including constant stiffness against disturbances. How to implement “evolutionary adaptation”, human acting as the “agent of evolution”, then?

Neocybernetic adaptation principles are simple, in principle, and can readily be implemented also in real systems. There are relations to traditional control approaches: Applying the cybernetic view of semantics (together with the “snapshot”, its derivative is needed among the measurement data; see chapter 7) it turns out that *multivariate PD controls* can be implemented; there are also connections to *internal model control*. High dimensionality and noise could assumedly be tackled with in unstructured environments ... This sounds like a panacea, and such general solutions probably never exist. Perhaps one should look at the cybernetic models more like methods towards implementing sophisticated data mining and process monitoring, perhaps better matching and supporting the mental views of human experts than what the traditional statistical tools can do (see discussions in chapter 7). The automated “human-like” preprocessing of the huge bodies of the measurement data and historical time series, finding relevant correlation structures among signals, makes it possible for the human expert to explore and perhaps exploit the available information more efficiently.

When extending the idealized cybernetic studies to practical controls, there are many challenges. The key problem is that when trying to impose the cybernetic principles afterwards on top of the existing automation system, where the structures already have been differentiated hierarchically, and when there are predetermined information blockages within those structures (see chapter 6), one somehow has to “bootstrap” the cybernetic machinery. For example, the following issues become acute:

- In industrial plants, there are predetermined goals of the system what comes to the products and their quality. This does not match the self-organization idea, where the system adapts to match its environment; thus, the adaptation process needs to be somehow controlled.
- Related to that, the agents (controllers in the plant) are typically not homogeneous and identical, what has been assumed this far. In real plants, controllers are in different locations, and they are tuned to implement only their specific control tasks — the SISO approach should be

  extended into a MIMO.

In addition to the theoretical aspects, there are also more pragmatic ones: In practical control, there is need of speed. The control quality is measured in terms of real-time reactions, and there is no time to wait until the statistical balance:

- The basic problem in dynamic control is that the time structure cannot be ripped off, it is the signal transients that are to be controlled. The controllers should not only be a simple mirror of the environment; they should be mirrors *between the past and the future.*

- Related to the previous item, there are the causality issues — in a real system, pancausality cannot be assumed: For example, the past measurements cannot be altered by later-time feedbacks. To implement feedback through the environment, external structures are needed (see chapter 7).

One approach is presented in chapter 7, where the ideas of *biomimetic control* are discussed. it turns out that such approaches can be studied in the framework of *model predictive control,* where the model-based estimation of the future is tried to be regulated by applying appropriate actions in current time.

In any case, to implement the cybernetic adaptation, the system must be stable to begin with. The independence of the controls is an advantage, but it is also a disadvantage: *Stability of the overall system cannot be assured* during adaptation. This discourages all practicing engineers — before cybernetic control can become reality, further studies are necessary.

It needs to be recognized that control theory is not in all respects an appropriate framework to understand cybernetics as there are many practices that are in contrast with cybernetic intuitions. Indeed, control is seen as a stereotype of reductionistic engineering-like thinking, systems being localized and divided in separate blocks, and within them control being centralized. One should never underestimate the inertia that is caused by the role of practicing automation engineers and plant operators not willing to alter their practices. The plant-floor level constitutes yet another cybernetic (memetic) system with new sets of tensions. One can expect some of the counterarguments to be rather fierce: For example, a practicing engineer does not want to compromise the plant stability at any cost (there is a big difference here between the engineers and economists who are familiar with risks and complex environments — see next section).

5.3 Towards wider views

The presented ideas of information-oriented control-based perspective are so simple that some comments can be said in general also about truly complex systems without the knowledge of the details of the systems or their numeric parameters. It seems that the most complex of systems, the memetic ones, also share the behaviors that can be motivated more convincingly in better quantifiable environments.

5.3.1 "System cybernetization"

There are two ways to implement enhanced controls in a cybernetic system: Either the controls can be made more accurate, or the controls can be made faster. These objectives can be reached not only through making the model ever better, but specially by implementing tighter coupling. In a cybernetic system, extreme optimization results in "stiffness" of the system, and worsened fitness in changing conditions (see next section).

There are many details in the control structure that can be manipulated to enhance the control — in complex cybernetic systems, the model adaptations can be more complex than in typical adaptive controls, as the system structure also can change; the developments can even take place in separate systems, and in different phenospheres. The structure changes are again related to processing of the critical substance, information: Either enhanced capture, transfer, or usage of this information. When speaking of memetic systems themselves processing information, the critical resource is actually *knowledge,* or "knowhow" about clever usage of the available information. An intelligent agent constructing such a system is always "at the edge of understanding". For example, constructivistic systems (technical, scientific, ...) evolve so that as soon as there is some new understanding about relationships among variables, it is exploited to increase system performance (if there are no compensating drifts, like cost, etc.). This becomes manifested in industrial plants, for example, where new controls are introduced to compensate deviations from the reference values if some new relevant measurements are available, thus making the system remain better in balance — and become more cybernetic. Otherwise there is assumedly evolutionary disadvantage, as the system is "less cybernetic" than it could be. These developments are implemented by humans, but, after all, the system follows its own evolution where individual human signal carriers have little to say.

The cybernetization developments have to be gradual, as the world changes in unpredictable ways as changes in the structures are employed. A clever balance of opposing needs (tensions) cannot easily be determined by a centralized mastermind — if some specific aspect is omitted, all vacuums will be filled somehow through unintended developments. Also the development efforts must be cybernetically balanced. Perhaps the best example is the downfall of the late Soviet Union, where the goal assumedly was to reach a better society — by applying the cybernetic governmental steering following the best theories of centralized control. However, the means and ends were not in balance as they were centrally controlled. Again, the main problems in Soviet can be characterized in terms of information extraction and exploitation: In data input, there were problems as the statistics were forged and not accurate; information was available too seldom in the five-year plan frameworks; information transfer (specially in the low level) was defective because of censorship and scarcity of communication devices; and, finally, when the controls were applied, they could not be enforced because of decline in moral standards — this decline also being caused by ignoring the sophisticated cybernetic balances in social and ethical systems.

So, complex systems seem to develop autonomously towards becoming more and more cybernetic, as being led by a guiding hand (see chapter 9). Regardless of the domain, the limiting factor in this evolutionary process seems to be related to extracting and exploiting information (or knowledge). Typical examples are found in todays working life. First, study the other prerequisite for "cybernetization " — better understanding of the system and gaining more information. This is implemented through supervision, questionnaires, and more paper work in general. And the other prerequisite — applying more efficient controls based on the acquired information — is implemented through increasing administration, organizational changes, etc. This all is introduced in disguise: Who could object to "missions and visions" or "developmental discussions"?

Speaking of terminologies: The system of language use is an interesting example of cybernetization in memetic systems. It seems that as the culture proceeds towards its stagnation, it is a comprehensive decline: For example, when the language becomes more "civilized", certain ways of speaking become obsolete and are substituted with bureaucratic, politically correct ways of speaking. However, small talk with mere cumulating periphrases becomes void, there is loss of dynamics when the variations are eliminated in the well-balanced refined utterances. When concepts lose real content, they are less capable of capturing the "flesh and blood" — and the mental constructs can only receive their meaning through interaction with the brutal reality. As discussed in chapter 7, true understanding goes only through two-way interaction with the environment. — It seems that there exist languages (like Finnish!) where the dynamic range still extends from very fine nuances to extreme bursts, concepts being clear and accurate, but still poetically open-ended. Surprisingly, perhaps it is such "less cultivated", least cybernetized languages that are best suited for expressing oneself — or for doing science, explicating and perceiving the real world outside our standard constructions?

The result of system cybernetization is that diversity becomes eliminated. What happens when finally all degrees of freedom vanish?

5.3.2 Faith of systems

It seems that all development ends in a collapse. If a system of cybernetic systems are let to adapt freely, catastrophes are unavoidable. How to control the adaptive control without paralyzing the system altogether? — at least, Nature has not found the way to do this. One cannot backtrack from a dead-end, after evolution there is a revolution — again see Fig. 5.5 (another perspective to "saltationism" is studied later in chapter 7).

The mathematically oriented *catastrophe theory* flourished together with chaos theory back in 1980's, trying to explain the processes beyond collapses. The goal was to understand continuous mathematical structures that give rise to abrupt behaviors: Why the once stable balances finally become unstable. However, the trivial one-function experiments did not have very much connection to real-life. In the framework of cybernetic systems one can now qualitatively understand such processes with no additional fancy theories: The key point is (again) the nonideal structure of information acquisition, and the resulting hierarchic structure of systems in different time scales.

Above, in Section (5.2.2), it was observed that a cybernetic adaptation strategy does not necessarily collapse — is there not a contradiction? — There is *not,* because now one is studying wider perspectives: In (5.2.2), it was assumed that the environment remains stationary, whereas now *structural changes* in the environment and in the system itself have to be taken into account — after all, true evolution is change in structures, not tuning of the parameters within existing structures. A closer analysis reveals that there are internal and external reasons for catastrophes. The internal reasons can be seen to be caused by the fast-scale structures changing, and the external reasons are caused by the slow-scale ones.

The fastest, catastrophe-like changes in the system balance can be explained in terms of nonlinearities — gradual changes in the system finally push the system onto the watershed boundary, and after that a new attractor is suddenly found. Such behaviors can easily be explained in the framework of sparse-coded nonlinearities, where some degrees of freedom can remain latent and completely inactive until the conditions are favorable (see chapter 6). As the history of memetic developments reveals, new ideas can remain ignored for a long time — after the turning point, developments can be very abrupt. Individuals are, after all, just noise when looking at the cybernetic systems that are based on statistical models, and developments can become relevant only after the whole population is ready to employ them. There is no evolutionary benefit if too smart enhancements are introduced too early — the key point is that the ideas remain available in the systemic memory (genome, or "menome").

The evolutionary changes within a system can often be characterized in terms of increasing coupling, or the parameters q_i increasing, finally the enhancements ending in structural changes. Flourishing systems are living *at the edge of chaos,* trying to capture the most up-to-date information (or knowledge); however, beyond the borderline determined by the information bandwidth, the visible variation is mostly noise, and the once acquired structure will be lost. What is then the appropriate frequency limit? The system guessing right wins it all. Explicit optimization is not easy here. For example, when making controls faster, the continuous processes typically become discontinuous at some stage as the acquisition of information cannot be immediate. And such discrete-time control systems behave in very different ways as the originally assumed continuous ones: As the sampling rate becomes too fast as compared to the system dynamics, increase in the noise sensitivity follows, and robustness is challenged in changing long-term conditions. There are real-life examples of such tendencies: For example, in "quartal capitalism" samples are taken and controls applied every 1/4 of the year, even though the market dynamics has the range of years; also in modern politics, long-term planning becomes impossible as the politicians have to take care of their everyday popularity according to the population polls — and, what is more, the real time constants in a society can be decades! In both cases, too fast adaptation and control actions can lead to loss of informative excitation and problems with stability.

The structural impacts coming from outside, or from the environment, are caused by low-frequency phenomena. Once some dependency structure that a system exploits has been visible for a (too) long time, it is probable that a slower system takes over that resource. The slowest processes are the most dominant in the long run, and the faster ones are left completely empty-handed, becoming unstable, the statistical balance corresponding to their local models being lost. When the universe gets older, ever slower dynamics become visible, and there is room for new systems to be born in the low-frequency end of the spectrum (again, see Fig. 5.3). When the behavior of the nominal state (or when the "fixed" environment, as seen by the faster system) changes, models for variations around that nominal states become outdated. Hierarchy of systems is like a tree, slower ones being nearer to the "root": When the "trunk" is adjusted, the "leaves" can be violently shaken. The overall system structure cannot change without making its subsystems outdated. Remaining fixed to protect its own fine structure would mean system stagnation.

The finer the constructions become, the larger are the catastrophes — this applies also to memetic systems. Indeed, the magnificent span of German philosophies during some 200 hundred years (ending in a complete catastrophe in 1945), starting from Immanuel Kant, continuing with Hegel himself, Arthur Schopenhauer, Karl Marx, and Friedrich Nietzsche, accompanied by the ideologies of Friedrich Schelling and Johann Fichte, and spiced by von Goethe and von Schiller, is itself an example of such ambitious mental endeavors that can only end in a *nemesis.* Indeed, it was Hegel himself who observed that the state of peace is stagnation, and war has positive moral value: One understands the "real values" again, there is *katharsis.* Along the same lines, the larger scale downfall of the entire culture was studied by Oswald Spengler. But the ideas are still there, the latent thoughts someday having an incarnation as some kind of a synthesis.

To avoid deadlocks of development, mechanisms of *regeneration* seem to be programmed deep in the structures of more sophisticated systems: The cycles of death and birth makes it possible to get back to a fresh start.

5.3.3 Coordination of catastrophes

When this dual nature of balances and catastrophes seems to be such a natural part of cybernetic systems, perhaps it cannot be all bad?

The unavoidable fact is that all complex enough environments are changing over time. One reason for this is that the environments are composed of coevolving systems, and these processes never reach the final state — or, if you start waiting for that, you will be hopelessly late. This dynamic nature of the world is general, it can never be escaped by any system, and it applies fractally in all scales; again, according to Heraclitus, "panta rhei". There is a vicious circle here: World evolves as the systems evolve, and as the world evolves, systems need to evolve. What is more, such changes are not only quantitative — when they continue long enough, quantitative becomes qualitative, and the whole system structure becomes outdated. This is typical in evolutionary systems.

To implement up-to-date control of their environments, and to survive in competition, the systems have to constantly update their models of the environments. Only change exists, but, according to the neocybernetic principles, balances are to be modeled. It seems that nature has found a practical way to gather accurate balance information even in changing environments: It seems that in some sense nature "discretizes" the time-variant processes, so that the processes take place in discrete time rather than in continuous time. First the environment is frozen, then a snapshot is taken, and as the internal tensions cumulate, suddenly the tensions are released to burst the old structures to have a fresh start. During the balance periods optimization of parameters within the structural framework takes place, applying the smooth neocybernetic adaptation strategies, but during the collapses, new structures are introduced to escape the local minima. Truly, the catastrophes themselves do not deliver information, they only produce noise and chaos: It is the balance periods between the catastrophes that are the cookers of information. Catastrophes on the lower level are crucial for the well-being on the higher level to reset the information-producing lower-level systems so that fresh information becomes available. The higher-level system is a model over the possible solutions on the lower level.

How can all this be explained — this all sounds very purposeful: It seems that one needs external control to coordinate the actions, to initialize the system, to run the processes, to collect the data, and to exploit the information. Can the above scheme be seen as more than a metaphor? Again, no master mind is needed to orchestrate the alternation of the "sample and hold". It just seems that "perfect control" — the property of the ultimate survivor in evolution — is an internal contradiction, resulting in extreme sensitivity and eventual collapse of the system. This is the nature's mechanism to guarantee the evolution and emergence of ever higher-order systems; at least, when looking back from the higher level, all lower levels have been obeying the this principle. In a way, nature has built this "apoptosis", or programmed death, in all its systems. And it seems there is automatic synchronization: Only after the properties of the environment are mapped, the controls can become complete — and, after that, it is the whole construction becomes unstable at the same time. Overall stagnation can be reached only when all subsystems have found their models, and when a collapse is then launched at some location, because of however small disturbance, the disturbance soon escalates, wiping away all submodels at the same time.

Even though the continuous processes become discretized, there is no one-to-one coupling to the time variable, and the strong tools from discrete-time dynamic system theory are not available. If trying to model the succession of catastrophes and balances, it is the transitions that are relevant, no matter when they happen, and modeling tools for *event-based system* could be applied. Unfortunately, there exist no strong analysis tools for such systems.

Can anything be said about the catastrophes in general? It is evident that individual processes, or unique catastrophes, cannot be individually modeled — but if seems that the catastrophes are by no means unique, they seem to repeat all over again. One can perhaps abstract over individual catastrophes and find a model for them on some slower time scale.

if the lower-level cycles of catastrophes and balances are correlated, it is information to be utilized. The only problem here is that for the most interesting systems, one cannot see the big picture yet, as one is living in the middle of the turmoil and perhaps emerging new order. Whenever the higher-level structure can be seen, it already exists, and our predicting attempts are late. It seems that such behaviors can be analyzed only in retrospect. As a (very) crude approximation, it seems that there is some general constant here: For a system to sufficiently develop, there are about a dozen regeneration stages at the lower level: How many times cells renew during the lifetime of an organism on average; how many generations there are during a life span of a human culture; how many individual species get extinct before the whole ecosystem collapses. However, the variations here are huge: Some species just seem to be less vulnerable to the changing environmental conditions — but, in many cases, such relics seem to be secondary what comes to the main developments in the larger-scale system.

In evolution biology, there are mysteries: The developments in natural evolution seem rather peculiar. One of the questions is *where are the missing links.* It seems that there have been very different kinds of species following each other with no transitional forms; similarly on the ecology level, there was the era of dinosaurs followed by mammals, etc. A species can be there with no changes for millions of years just to be suddenly substituted. This kind of succession of balances and transients is known as *saltationism.* The lack of continuity in evolutionary processes has been used also as an evidence by creationists; however, as discussed above, this kind of behavior of bursts and balances assumedly is characteristic to all evolving cybernetic systems.

5.3.4 Beyond the balances

Balance is needed for healthy functioning of a system, but catastrophes are needed for healthy functioning of a "supersystem". There must exist variation on the lower level, otherwise higher-level developments cease. It would seem that it is the higher-level system that is running experiments on the lower levels, pushing those systems over their limit on purpose — but, again, there is no such master mind. Catastrophes are built in the cybernetic systems themselves, no matter if the generated excitation is ever exploited, or if it remains just noise in the universe. A healthy evolving system follows its *elan vital* until the edge of chaos — and beyond.

In some environments collapses in different scales are commonplace and — as it seems — generally accepted as unavoidable. The *stock market* is a great equalizer of tensions in economy, tensions manifested through sell and purchase prices, being a simple example where the balances should be found according to cybernetic principles. Again, the stock market dynamics is too fast as compared to real market dynamics: Analysts use their mental models reflecting the common beliefs, making the unquantifiable aspirations visible; these beliefs can be very volatile. The agents try to be smart, trying to predict the competitors and market reactions, thus making the stock market a constructivistic system that lives a life of its own, detached from the reality. The money is not necessarily where the needs are: The challenge of a modern society is to match these tensions — needs and means — and it is here where more cybernetic thinking would be needed, more sophisticated models of the interdependencies and their balances, not straightforward centrally-controlled legislation. In any case, it seems that the minor everyday catastrophes are, as seen from outside, only the mechanism of introducing the necessary excitation and information in the market — but inevitably "the big one" also comes some day.

Extending the observations in chapter 4, it can be claimed that a democratic society — if accompanied by transparency — is the most efficient political system in terms of information exploitation. It combines gathering of bottom-up agent-based innovations, and delivers top-down regulatory directives. But to remain "alive", perhaps democracy, too, needs its enemies, or some excitation from outside. The key question is how can the regeneration of the social structures be implemented in the "postmodern" society, where all destructive developments are prevented. Today, it is interesting to see what the alternative is — how long can Europe become older?

Also in natural everyday systems the "catastrophes" are a part of normal behaviors in healthy systems: The limits are being tested all the time. Without pushing the limits, the dynamic range becomes narrower. For example, take the living body: If the machinery is not "calibrated", if there are not the necessary degrees of freedom visible in data, missing compensation capacity against certain excitations is developed. If the body does not get acquainted with microbia, there can be an increase in autoimmune diseases. The genes only determine the gross structure, but the fine-tuning of the system is found as a interaction process with the environment: Diseases are minor catastrophes, extreme cases that determine the dynamic range of the system. And as they say in the United States: You cannot know what the business is all about before you have experienced some bankruptcies.

It is all cybernetic subsystems that are hungry for information: In extreme balance the system starves. This can be extended even to analysis of mental sanity: One needs to have "highs and lows" to experience what life is about. And extreme feelings seem to be the seed for higher-level memetic systems, at least what comes to artistic creativity. Of course, diseases are related to the loss of balance in the biological environment, mental diseases are related to loss of balance in the cognitive environment, and "social diseases" are related to loss of balance in the social environment. If there are no real political issues in a welfare society, the system becomes — concretely — insane. But extreme emphasis on the balance is a fallacy: If there are no real obstacles or problems, these will be imagined — or when life is too easy a healthy mind actively searches for challenges, to find the balance of feelings between danger and security. The "real artists" simply need to experience the highs and lows. Without mental explorations and excitations one has an incomplete model of oneself and of the world. This is where neocybernetics goes even beyond the Eastern wisdom: The goal is not extreme harmony or elimination of variation — as they say it, "in Hell you have merrier company". Such discussions can be extended even to purposeful life and what happiness is about: It is mastery of one's life, or awareness of one's capability of coping with all possible challenges one might face.

It has been observed that evolving morality, etc., are becoming fields of scientific study [83]. This is true, but there is another tendency, too: In the neocybernetic framework all biology is coming back towards more abstract philosophies.

It is tempting to draw some bold conclusions concerning issues that by no means have been seen as subject to scientific study. For example, *why there is evil, why there is poverty in the world,* or, *why there is suffering?* Indeed, suffering seems to be necessary for a cybernetic system to fully develop. There are two ends in the continuum – always somebody is the poorest. If there were no differences, the *heat death* would have been reached. Questions like *why there is death* can also be attacked: Death is dropping out from the dynamic equilibrium to the static balance, it is nature's means to assure regeneration in the system. Whereas death is the final catastrophe from the individual's point of view, it is necessary from the point of view of the wider-scale system. At some stage of the higher-level development, lower-level models are so outdated that it is easiest to start all over again.

Is it perhaps so that engineering disciplines, like understanding of control engineering, can give some mental building blocks for understanding of, for example, what *good life* is? What is more, it is not only ethics, but also other branches of philosophy that can be affected by the cybernetic considerations. These issues are studied in more detail in chapter 10. — However, next it is time to go back into details: It is there where the beauty is.

Part II

Further Studies and Intuitions

Level 6

Structures of Information beyond *Differentiation*

The key concept in cybernetic systems is information, availability of information determining the models that are constructed. Assumption of environment-orientedness means that it is the information coming from the environment that dictates the results in a more or less unique way.

Despite the assumed uniqueness there still are many ways how the world can be seen and how this view can be interpreted. As the neocybernetic models are based on observed correlation structures, by appropriate scaling of the variables one can implement continuous modifications to the information that is visible to the system. This all is familiar from principal component analysis. However, here the goal is to extend beyond the existent intuitions: What happens when the amount of available information increases? How can the *emergence of structures* be understood?

6.1 Towards more and more information

Being based on principal components, neocybernetic model is robust against high dimensionality. To assure maximum information availability, a reasonable strategy is to *include all available data among the measurements* — the modeling machinery can automatically select the relevant pieces of information. When the data dimension becomes high, there are also qualitative and theoretical benefits.

6.1.1 About optimality and linearity

Thinking holistically is a comprehensive challenge. For example, one should no assume that there is some centralized optimization criterion being reached for by the system. But if the data dimension is high enough, a common goal is a useful abstraction: It turns out that *optimality become reducible.*

The most straightforward way to assure the supply of information is to inflate the space of input variables, so that m, the dimension of input data, grows. To analyze this issue, assume that the cost criterion can be locally decomposed so that its differential change can be expressed as a sum of N weighted parts:

$$\delta J = c_1 \delta J_1 + \cdots + c_N \delta J_N \tag{6.1}$$

Here the sub-criteria are assumed to be locally linearizable, so that

$$\delta J_i = Q_i^T \, \delta u \tag{6.2}$$

for some parameter vector Q_i and variable vector u. If the sub-criteria are independent, for high number of variables there holds for correlations among different vectors i and j

$$\frac{Q_i^T \, Q_j}{m} \quad \to \quad 0, \qquad \text{as} \qquad m \quad \to \quad \infty. \tag{6.3}$$

The more there exist variables, the better random vectors become orthogonal. When solving for gradients, one has

$$\frac{\delta J_i}{\delta u} = Q_i \tag{6.4}$$

so that

$$\frac{\delta J}{\delta u} = c_1 Q_1 + \cdots + c_N Q_N. \tag{6.5}$$

Now, assuming that the variables are adapted along the negative gradient of some sub-criterion, so that $\Delta u = -\gamma Q_i$, the global criterion also goes down:

$$\Delta u^T \frac{\delta J}{\delta u} = -\gamma Q_i^T \left(c_1 Q_1 + \cdots + c_N Q_N \right) \approx -c_i \, Q_i^T Q_i < 0. \tag{6.6}$$

This means that if the sub-criteria are mutually independent, and if the input data dimension is high enough, the task of multi-objective optimization can be decomposed. Local optimizations result in global optimization.

What is more, when the data dimension is high, getting stuck in local minima is less probable. Multiple variables typically mean better continuity in the data space, and perhaps also evolutionary processes can be characterized in terms of "generalized diffusion". How about the cost criterion (6.1) then — is it not unrealistic to assume linear additivity of the sub-criteria? Again, it is high dimensionality that helps to avoid problems. The more there are features (variables) available, the more probable it is that the problem becomes more linear (compare to the idea of *Support Vector Machines,* where a complex classification problem is changed into a simple problem in high dimension).

6.1.2 New sensors and innovations

When trying to affect the modeling results, selection of variables to be included in input data is the most important decision. How to assure high dimensionality and fresh information in the data, where to find the new sources of observations?

New innovations and new sensors are needed by the system — the term "sensor" being used in a relaxed sense here, as the information capture is to be seen in the holistic perspective. It does not matter what is the physical manifestation of the sensors, as long as the acquired information can cumulate in the model structures. Some examples are given here.

- **Spatial distribution** can be utilized, that is, information from spatially distinct locations can be used. This far it has been assumed that a system is isolated — however, in a real ecosystem, neighboring systems are in close connection, and they can be modeled as a whole. Specially, assuming that there are no limitations for seeds to spread within some area (or no limitations for information flow), the spatial structure can be "collapsed", assuming that the spatial distribution delivers relevant material about the ecosystem in general. This can be utilized when constructing the covariance matrices: Plentiful fresh data and variation is available when each subregion within the ecosystem delivers its contribution to the behaviors of the environmental variables.

- **Temporal distribution** can also be utilized, that is, information from temporally distinct time points can be used. Assuming that a species in an ecosystem has some (hard-coded) memory, it is not only the current state of the environment that is seen by the population, but also the time history: If the previous year was bad, the population is lower this year, no matter what are the current circumstances. The longer-living the individuals of a population are, the longer is the "memory", too. When cybernetic models are constructed for such time-series data, it is no more simple PCA that is being carried out; it is dynamic modeling in the framework of (implicit) *subspace identification* [60]. It can be assumed that if the food level variations are low, then — after adaptation — the environment seems to support longer-living species. Is it because of this that predators live longer than prey, the information being filtered more on the higher trophic layers?

It turns out that the more there are variables, and the more there are possible variable combinations — and the more there are ways to select the "interesting" or most relevant features, different selections resulting in different models and different views of the world. This is a special challenge in constructivistic systems, where the space of candidate variables is potentially infinite; in psychology, one speaks of the *Barnum effect,* meaning that when there are enough degrees of freedom, any model can be matched against the data (making numerological studies, for example, often astonishing).

6.1.3 Example: Transformations implemented by nature

Frequency domain was employed in the previous chapter to study information distribution among subsystems. But such considerations are applicable not only at the ecosystem level — it seems that also within a single individual similar analyses are appropriate. Specially in the processing of auditory and visual information clever data preprocessing is needed to extract fresh features from the temporal and spatial data. Again, it is a nice coincidence that there are powerful mathematical tools available for analysis and understanding of such features.

When auditory, time-domain signals are received, the cilia in the inner ear implement spectral analysis: Depending on the frequency, sound waves can penetrate different distances in the cochlea. As the cilia are connected to the auditory cortex, energy in each frequency range becomes an input signal of its own, the number of inputs thus becoming expanded. What the brain then can see in the preprocessed signals is combinations of formants; this means that the patterns being modeled are *phonemes.*

It seems that similar frequency-domain reconstruction of signals takes place also when visual signals are processed; however, now the information is not distributed temporally but spatially. Simple networks of neurons can implement (two-dimensional) discrete Fourier transform. This kind of coding of the images is beneficial because cross-correlation between two transformed images efficiently reveals the dislocations and structured differences among the images. For example, movements within the field of vision are manifested when successive transformed images are compared; on the other hand, depth cues become available when using image pairs acquired from nonidentical locations (from the two eyes). The succession of parallel / sequential transformed image vectors is interpreted as input data samples; when the correlation structures among data are modeled in the neocybernetic spirit, the resulting sparse components (see later) perhaps reveal natural-like decomposition of visual patterns. This kind of extra information concerning spatial dependencies among visual entities can perhaps explain the properties of three-dimensional vision.

6.2 Blockages of information

When there is plenty of data available, not all need to be used. Here, some examples are given how the results can be controlled by explicit channeling of information, by explicitly determining structures of data flow. In a sense, it is all about implementing non-idealities again — the ranges of seeing information are limited.

6.2.1 Hierarchic models

As an example, study a cybernetic system with the following model matrices (assume "clever agents"):

$$A = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \cdot & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \end{pmatrix}. \tag{6.7}$$

Dots in the structures mean that those connections are non-zero, whereas empty slots denote missing connection. The above B matrix form can be appropriate in sensor/actuator structures, where each actor has its own measurements. There is no complete information available, and data flow becomes localized. The nonideal flow of information introduces distortion in the data, and the analyses in chapter 3 become outdated: The degrees of freedom in input data are no more a limiting factor, non-trivial structure emerges even though $n = m$. Closer analysis reveals that the basis vector ϕ_i is dominated by the local measurements u_i.

More interesting results are found if one has triangular interaction matrix, each actor only seeing the actors in front of it, the last actor being capable of seeing all information. The structure becomes strictly hierarchic:

$$A = \begin{pmatrix} \cdot & & \\ \cdot & \cdot & \\ \cdot & \cdot & \cdot \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{6.8}$$

This means that the first variable is not affected by the other ones — it lives a life of its own, exploiting all the information that is available in input. Thus, it alone constructs a principal subspace model of the data; because this model is one-dimensional, it basis vector must coincide with the first principal component axis. In this sense, the first variable implements (trivial) principal component analysis rather than principal subspace analysis. Such reasoning can be recursively continued: The second variable is affected only by the first variable representing the first principal component, meaning that its contribution is deflated from the data. This way, looking at the second variable alone, it is the *second* principal component that must be represented by it. And the same analysis can be continued until the variable n, meaning that the hierarchic structure implements explicit principal component analysis. Because of the information blockages, principal components get separated, and structure emerges.

6.2.2 "Clever agent algorithm"

Implementing an algorithm is a compromise between theoretical and practical aspects. Now it seems that the nonideality — triangular blockage of information, as motivated above — enhances convergence, as the variables disturb each other less. It turns out that the Hebbian/anti-Hebbian adaptation becomes a useful PCA algorithm, as it is robust — there are few free parameters — and because the explicit construction of the covariance matrix $\mathrm{E}\{uu^T\}$ is avoided: In the cybernetic cases, m is typically high, and the covariance matrix can be huge.

In `Matlab` syntax one can write the "vanilla" algorithm as showqn in Fig. 6.2 (matrix `U` containing the k sample vectors u^T as rows, and matrix `Xbar` containing the k internal variable vectors $\bar{x}^T$ as rows).

The data structures `Exx` and `Exu` are initialized to small values (matrix `Exx` having to remain positive definite at any time). The parameter λ determines the adaptation rate. After convergence, the basis vectors can be picked out from the matrix $\phi^T = \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}u^T\}$.

As an example, a case of coding hand-written digits is represented. As data material, there were over 8000 samples of handwritten digits (see Fig. 6.1) written in a grid of 32×32 intensity values [50]. The 1024-dimensional intensity vectors were used as data u, and the algorithm was iterated until convergence. The results are shown in Fig. 6.3.

Figure 6.1: Examples of handwritten digits

```
while ITERATE

  % Balance of latent variables
  Xbar = U * (inv(Exx)*Exu)';

  % Model adaptation
  Exu = lambda*Exu + (1-lambda)*Xbar'*U/k;
  Exx = lambda*Exx + (1-lambda)*Xbar'*Xbar/k;

  % PCA rather than PSA
  Exx = tril(ones(n,n)).*Exx;

end
```

Figure 6.2: **Algorithm 1:** Hebbian/anti-Hebbian PCA by "intelligent agents"

6.2.3 On-line selection of information

There are information flows and blockages on many levels in an adapting system, and frequency-domain characterizations are possible here, too. The slowest-scale control of information takes place in the adaptation processes: For example, gene pools that restrict information to remain within the species implement en extreme block for spreading of information. The results become visible as peculiar evolutionary developments on the species level.

In the other extreme end, the information blockages can also be very temporary. For example, the routing of information can be dependent of the actual signal properties — meaning that the signal path is nonlinear. As seen from the opposite point of view, it can be said that nonlinearities are information filters.

Linearity means homogeneity and predictability, whereas nonlinearity is the key to emerging differentiation among structures. When dropping the assumption of linearity, the strong guidelines are lost: There is an infinite number of possible nonlinearities available, and there is no general theory to understand the

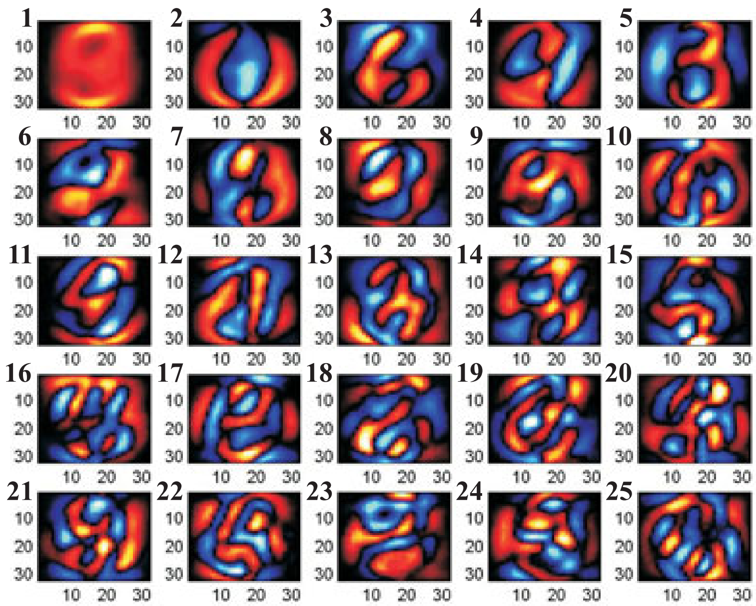


Figure 6.3: The 25 principal components extracted from the handwritten digits. The 1024 dimensional feature vectors have been decoded as planar patterns to reveal their connection to input data properties (dark regions mean that there is no correlation with the feature and the corresponding pixels; light blue regions denote high negative correlation, and light red denote high positive correlation). Because of the hierarchically structured feature extraction, the sparse subspace has been decomposed into the actual PCA basis vectors: First, there is the mean vector, and thereafter the correlation structures are presented in the mathematically motivated decreasing order. The coding is efficient when there is scarcity of latent variables, but the physical relevance of the features is questionable when the basis dimension becomes large

resulting functionalities. What kind of nonlinearity to choose, then? It turns out a good compromise is a function that implements a volatile *switch.*

$$f_{\text{cut},\text{i}}(x) = \begin{cases} x_i, & \text{if } x_i > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{6.9}$$

This *cut function* (see Fig. 6.4) lets positive signals go directly through, but eliminates negative ones. This function is piecewise linear — this offers theoretical benefits as between the transition regions linear model structures are applicable. There exist also strong physical motivations for this selection of nonlinearity: Whatever are the signal carriers — concentrations, frequencies, agent activities — such activities can never become negative. In more complex cases, for example, when modeling gene activation, the cut function is still applicable: Remember that there are excitatory and inhibitory transcription factors

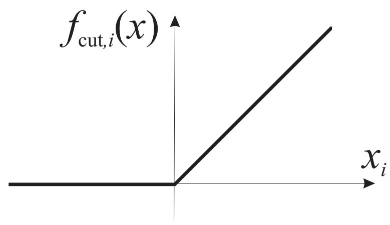


Figure 6.4: Nonlinearity for online-blocking of information

controlling the process; there must be excess of excitation to start the process in the first place, and the more there is excess, the more chromatin packing of DNA opens up to promote gene expression.

Such a simple form of nonlinearity makes it possible to implement "soft" transients between structures. When a variable becomes active, a new dimension in the data space becomes visible. As the nonlinearity is monotonic and (mostly) smooth, optimization in pattern matching can thus take place among structures.

As explained in [92], locally unstable models become possible because of the nonlinearity: Extreme growth in variables is limited by the cut functionality. When combined with a dynamic model, it is possible to implement bistable "flip-flops", where minor differences in initial states or in the environment result in completely differing outcomes. When comparing to natural systems, only the stem cells are assumedly free of such imprinting; in practice, the evolved "epigenetic states" can be very stable after such a development has started (for the coloring of animal fur, see [81]). These peculiarities that are made possible by nonlinear structures are not elaborated on here; the cut nonlinearity will be employed in what follows only to *boost* linearity.

6.2.4 Switches and flip-flops

To see how the nonlinearity can affect the originally linear and well-understood smooth behaviors, an example is needed. Assume that the "cut" function is included in the system model so that one has

$$\frac{d\,\xi}{d\,t}(t) = A\,x(t) + B\,u, \tag{6.10}$$

where the visible activities are limited to positive values:

$$x(t) = f_{\mathrm{cut}}(\xi(t)). \tag{6.11}$$

Applying this model structure, a "comparator system" was simulated with two mutually inhibitory subsystems:

$$\begin{pmatrix} \dot{\xi}_1(t) \\ \dot{\xi}_2(t) \end{pmatrix} = \begin{pmatrix} -\gamma_1 & -1 \\ -1 & -\gamma_2 \end{pmatrix} \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}. \tag{6.12}$$

The negative non-diagonal elements in A matrix implement negative feedback among the subsystems. In simulations, $\gamma_1 = \gamma_2 = 0.75$; this means that the eigenvalues of the matrix A are $\lambda_{1,2} = \gamma_{1,2} \pm 1$ or $\lambda_1 = 1.75$ and $\lambda_2 = -0.25$,

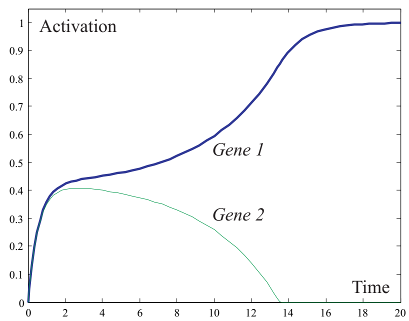


Figure 6.5: Incoming concentration ratio $u_1/u_2 = 1.00/0.99$

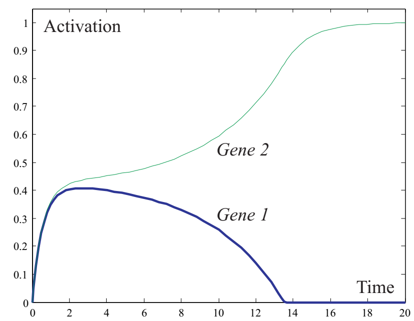


Figure 6.6: Incoming concentration ratio $u_1/u_2 = 0.99/1.00$

and the linear system without the additional nonlinearity would be unstable, x_1 and x_2 escaping to infinity, the one in positive and the other in negative direction. The variables escaping in opposite directions "pump" each other; as the nonlinearity prevents variables from escaping in negative direction, it simultaneously stabilizes the positive variable as well.

The simulation results (starting from zero initial values) are shown in Figs. 6.5 and 6.6. It seems that in this framework inhibition and excitation together define a system where some variables stabilize to non-zero values and other to zeroes ("winner-take-all"), depending on the input value distribution: Using the above model, x_1 wins and x_2 vanishes altogether if $u_1 > u_2$, and vice versa, the inputs being constant. It turns out that, qualitatively, the behavior is rather robust regardless of the exact parameter values.

The presented model structure makes it possible, for example, to define a genetic functional "state". Remember that the gene expression is controlled by specific inhibitory and excitatory transcription factors, these transcription factors forming a complex network, all of them being products of the activity of other genes. Minor changes in input concentrations make the resulting environment within the cell completely different: The "flip-flops" take either of the alternative values depending of the ratio between inputs, and once they have ended in some state, it is difficult to change it. In this sense, associations to properties of stem cells are easily made: A cell that has specialized cannot any more take some other role. Other bonus intuitions are also available: Today, there is the link missing between strictly biophysical considerations and qualitative ones. The purely numeric, quantitative, continuous approaches and the qualitative and discontinuous approaches are incompatible. The claim here is that the presented model makes it possible to study *emergence of structures.*

In the framework that is boosted with nonlinearity, competition among agents can be intensified: Effects of substructures can be wiped away altogether. Such extreme behavior is only possible in nonlinear systems because it is due to the nonlinearities that the system remains stable. Indeed, there emerge alternative minima depending on the initial values. In complex cybernetic systems, mastering such local solutions is of utmost importance, and rather than studying individual nonlinearities, a higher-level view is needed.

6.3 Real world of nonlinearity

The basic neocybernetic model is linear. This was a reasonable starting point as there is no strong theory for nonlinear systems. Linear structures are, after all, always easy, as there exist unique minima in a given environment; for nonlinear systems this does not hold, and typically the results are not identical even if the environment remains the same. However, nonlinearity is the nature of the real world, and when the objective is to model it, the modeling machineries have to accept this fact. So, how to characterize nonlinearities, and, specially, how have the real systems managed to do that?

6.3.1 What is relevant, what is reasonable

This far, the method has been the starting point, and its properties have been examined; however, now concentrate on the applications. Now there is the whole wide world ahead of us, the class of nonlinear functions being infinite and indefinite, and one should be careful not to open the *Pandora's box.*

To have a balanced view of the problem and the possible ways to attack it, one can utilize the above discussions, and *exploit the cybernetic model of an existing memetic system:* Ideas have been competing in bright minds, and an equilibrium can be observed. In the spirit of the Delphi method [25], different arguments have been thoroughly discussed by experts — in the field of *artificial neural networks* the problems of capturing "natural data" have been intensively studied from different points of view (for example, see [36]). This ANN research is a well-established branch where compromises have been found between what seems promising from the point of view of representing natural data and what seems possible and practical from the point of view of available tools, and today, a "model of models" can be compiled: What are the dimensions of the problem, what are the interesting applications and promising methodologies. Within such a memetic "supersystem" the degrees of memetic freedom are manifested, helping to see the "intra-paradigm paradigms", combinations of aspirations and visions, where different points of view are weighted in different ways.

Some advantages can compensate other disadvantages. For example, despite the theoretical deficiencies, there is so much physiological and mathematical support for linear structures that today there exists a large body of literature, and still there is active research in that direction. As a paradigmatic approach, there are different kinds of algorithms to implement principal component analysis, and, further, there are different kinds of extensions to the basic models (see [26]). These studies are motivated by the physiological studies concerning the Hebbian neurons, and they are further boosted by the strong theoretical intuitions and interesting applications: The research is still going strong specially in the field of *independent component analysis* [41].

Another family of intuitions have motivated the study of *feedforward perceptron networks:* It has been observed that within this model structure, all smooth nonlinearities can be approximated, at least in principle. In practice, this unlimited expressional power is a problem: To select among the alternative functional structures and to determine the parameters within the selected structure, there is need for very high numbers of data. Often some additional assumptions are (implicitly) employed for pragmatic reasons — for example, typically one limits the search to smooth mappings. There is also another class of model structures with equally high expressional power: The *radial basis function networks* are based on basis functions, simple prototype functions of localized support (like Gaussians); when such functions are scaled and scattered appropriately in the data space, their combined envelope can again be matched with any smooth function. As compared to the perceptron networks, the basis function networks are better manageable as the representation there is more local and easier to interpret.

The above ANN structures have continuous output, and they can be applied for function approximation; a more special application is *pattern recognition,* where one only needs discrete output. There is a very special network architecture that deserves to be mentioned here because of its close relation to the neocybernetic discussions concerning dynamic models and balances: In *Hopfield networks* the input is given as an initial state to the system, and a dynamic process searches for the minimum of the energy function, revealing the pattern that is nearest to the input. The construction of the network is such that it assures that the attractors of the dynamic process are the stored patterns. However, as compared to the neocybernetic model structure, now there is no input; the end result is unique after the initial state (the incomplete pattern to be completed) is given.

All of the above neural network structures are mathematically rather involved; in the other extreme, there are the intuition-oriented approaches where it is the actual brain structures and functionalities that one tries to reproduce. One of such intuitions concerns brain maps: The mental functions have their own locations in the brain, related functions and patterns assumedly being stored near each other. The *self-organizing maps* try to mimic the formation such (two-dimensional) maps [46]. There are many applications what comes to data visualization: On the SOM map the high-dimensional data distributions are often made better comprehensible. As the high-dimensional real-valued vectors are coded in terms of N integers (map nodes) only, there is extreme data compression, and information loss cannot be avoided. The most interesting issue about the SOM is that in some sense it seems to match our mental structures — perhaps there are lessons to be learned here (see chapter 7).

It seems that all ANN methods attack only one issue at a time. To address different needs, a compromise is needed; and it can be claimed that the neocybernetic model can be extended to combine the ideas of basis functions, dynamic attractors, and intuitive considerations, combining comprehensibility and expressional power in the same framework.

6.3.2 Models over local minima

For a moment, it is beneficial to look at modeling in the probabilistic perspective. When seen in the probabilistic framework, the goal of a model is to capture the data distribution, the model explaining as economically as possible where an individual data sample is located in the data space.

How can the neocybernetic model be characterized in terms of distributions? It is not the degrees of freedom alone (as studied in chapter 2) that would capture the variable distributions; when elasticity is also taken into account, tensions pulling the system towards balance, samples tend to become clustered around the nominal state. Assuming that the scores have normal distribution (being results of many independent equally distributed random variables being added together), and assuming that the basis axes are mutually independent, one could use the *multivariate normal (Gaussian) distribution spanned by the degrees of freedom* as representing the behaviors of cybernetic variables. However, natural data is *multimodal,* it cannot be represented by a single one-peaked (Gaussian) distribution — but an arbitrary smooth distribution can be approximated as a combination of (Gaussian) sub-distributions. Together the candidates define a basis, so that (if there is enough of them and they are appropriately combined) one can implement a *mixture model.* Strictly distinct clusters are implemented if the representation is *sparse* (see below).

Thus, the radial basis function metaphor would be applicable here; however, the structure also suggests more appropriate interpretations. Because the basis functions are now linear, the vectors ϕ_i determining the basis functions through the dot product operation, so that the matching against the input is calculated as $\phi_i^T u$, the basis functions have infinite support and there is no finite maximum. One does not only have a mixture of basis vectors that determine the distributions, one has "basis subspaces", determining the directional components present in data. The structural components define *feature axes* to be exploited by the higher-level model.

This far the model has been assumed linear. If the representation is *sparsely coded,* so that only a subset of features is employed at a time (see 6.4), the contributions of some features (the least significant ones) being cut to zero, there emerge structural alternatives, not all submodels sharing the same components. The sparse coded structure, where the substructures are linear, becomes *piecewise linear.* When the active components vary, there exists a wealth of candidate structures. Out of the n available features, in principle one can in the sparse coded case construct as many structurally differing distributions as there exist partitionings of the n variables between active and inactive ones. For large n this becomes a huge number. Such wealth of distributions is difficult to visualize: The sub-distributions are not clusters in distinct locations, and, indeed, one should not think using intuitions from low dimensions. What does this kind of a world look like, is elaborated on in chapter 7. In any case, such sharing of features is versatile, and it helps to reach generality and efficiency of coding; from now on such mixture of linear submodels is assumed as the prototypical model when the strictly linear models no more suffice.

When the mixture metaphor is employed as the basis of modeling, some extensions to the adopted model framework are needed; in a complex hierarchic system it is not only the highest level that is assumed to be nonlinear, but the model extension needs to be applied fractally. Before, the models were based on the linear features determined by vectors ϕ_i, and stacking of submodels was straightforward, linear structures being directly summable. Now one needs to extend to *nonlinear features:* When seen from above, the mixture model also defines a "feature" to be exploited by the higher-level model. To facilitate this, to make the extended model compatible with the linear model, the mixture model needs to look the same as the simple one, when seen from outside. The "interface" of the simple model is the latent variable activity, or score of the feature, in the form $\bar{x}_i = \phi_i^T u$; the submodel only delivers one real number to the outer world, revealing the match of the input data with the submodel. Also the mixture model has to be manifested in the similar manner to the outside world when such a model is further being used as a submodel — how to accomplish this?

The experience with the linear case helps here: The goal of the basic neocybernetic model is optimum match with the environment, and as complete reconstruction capability as possible so that no variation in the input data is lost. The latent variable $\bar{x}_i$ is a measure for how the submodel ϕ_i alone managed in this matching task, or, indeed, how much this submodel was "trusted" in this task, the balance of these latent variables being determined through competition among candidates. When ϕ_i are interpreted as basis functions, the outputs $\bar{x}_i$ represent the matches, or activities of individual, vector $\bar{x}$ revealing the success pattern, determining a coding of the prevailing environment. This view can directly be extended to the nonlinear case. The whole grid model is to be collapsed to one number characterizing the fit with the environment; let this number be called *fitness* of the model[1]. When employing the model, the cybernetic fitness criterion is how well the environment can be modeled, or reconstructed, and this can be expressed in the form $|\hat{u}|^2$, representing the length of the reconstructed input vector when the model is used for its reconstruction. To the outside world, the mixture model thus looks like

$$\bar{x}_i = \phi_i(\bar{u}) = |\hat{u}_i|, \tag{6.13}$$

where $\phi_i(\cdot)$ denotes some scalar-valued function, and $|\hat{u}_i|$ is the contribution of the submodel i in the input reconstruction, when various submodels compete in that task, and when equilibrium has been found. Remember that this "input reconstruction" actually means resource exploitation, making the assumptions about the same goals of subsystems generally justifiable. The value $|\bar{x}|$ becomes zero if the environment cannot be captured at all by the submodels, whereas if there is complete match, the whole variance of the input data is transferred further. It is also variance (average of the reconstruction vector length squared) that is a universal optimality measure in the nonlinear as well as in the linear case. Discussions concerning information, etc., thus remain valid also in the nonlinear case. The presented fitness definition abstracts away the implementation of the submodel, encapsulating it as a black box — indeed, it need not even be based on the presented mixture structure; there are no constraints as long as the model structure has mechanisms of producing the estimate $\hat{u}_i$. This means that the neocybernetic framework offers a general-purpose environment for studying very different kinds of coevolving complex systems.

No communication among submodels is needed: The model becomes balanced just in the same way as in the linear case. No matter how the individual submodels are implemented, they compete with each other, exploiting the resources; better models exhaust the available variation, leaving less resources for others to exploit. The coordination among submodels is again implemented implicitly through the environment, and there is no need for external supervision and "selection of the fittest" as in traditional clusters-based structures, etc. The universal "fitness criterion" is the modeling capacity: How well a (sub)model

[1] Indeed, there is a close connection to *genetic algorithms* here

can explain (and thus exploit) the environment. Trusting one's own observations, or the available remaining resources, makes it possible to implement local adaptation without compromising the emergence of higher-level structures.

To conclude, the mixture models constitute the basis functions for the next-level models. In the linear case it was the vectors ϕ_i that were thought of as characterizing the submodels, $\phi_i^T u$ giving the matches; in the nonlinear case it is $\phi_i(u)$ that returns the submodel activity.

The operation of the cybernetic model is defined through a dynamic process; similarly, the mixture model should be seen as implementing a *set* of such dynamic processes. Each of the submodels that determines a sub-distribution simultaneously determines an *attractor* in the data space, hosting a local minimum of the cost criterion, where the data matching process converges in favorable conditions. The final location of the fixed point within the basin of attraction is dependent of the input data. There are no "strange attractors" or the like, everything is quite traditional, being based on even (locally) linear processes and local balances. The proposed combination of linear and nonlinear structures seems to usually assure unique fixed points in the framework of many basins of attraction, thus combining simplicity and expressional power. However, being such a powerful framework, not very much can be said in general terms about such mixture models; one approach to examine the possibilities, based on simulation, is studied in chapter 8.

Model consisting of multiple attractors — this seems to be an appropriate way to model complex natural systems, too. Remember that nature is working in a distributed way, there is no central design unit: Finding the absolute optimum in a complex environment is just as difficult for nature as it is for humans. Nature is so varied because different solutions have ended in different local minima of the cost criterion. Perhaps a cybernetic model constructing a multiple model characterization over the alternatives better captures the natural diversity of natural systems? The cybernetic model can be seen as a *compressed model optimized over local candidate solutions.* This is a major difference as compared to traditional modeling where it is the only global optimum that is of interest. Remember that many problems of computability theory are concentrated on the NP-hard problems that are practically undecidable in large systems — but rather good local minima are easily found.

6.3.3 How nature does it

The mixture model seemingly has a complex hierarchic structure of submodels. Does such a "model library" need to be stored in some centralized location and maintained by some master mind? The answer, of course, is *no* — nature routinely runs such mixture models in a distributed manner.

Traditionally when deriving clustered basis function models, the key challenge is to determine the locations and the outlooks of the basis functions. Now it is the competitive learning among agents that carries out the matching against the environment in a distributed manner: The basis functions themselves are composed of still simpler basis functions — the agents themselves. When looking at cybernetic systems, it is important to recognize that it is not only one system that is running at a time: It is typically *populations* where there are individual more or less identical subsystems. This populations-based structure is quite universal, and it applies fractally to all levels of the systems: Within an ecosystem there is the large number of separate species, and within a species there are the individual animals[2]; in an economy there are the companies, and within the companies there are the humans; in a tissue there are the cells, and in the cells there are the chemical molecules.

Nature implements the whole mixture model in a parallel fashion, running the subsystems side by side, and constantly evaluating the performance of them. Optimization in such a structure is completely distributed. Each individual represents a local optimum having adapted to match its local view of the environment; the number of individuals representing a single solution reflects the relative goodness of the solution candidate, a good solution (or niche) being capable of supplying more resources to share. It is the whole set of functionalities that together characterize the nonlinear system of systems — the final mixture model representing a human, for example, being a coordinated-looking composition of the functions of its subsystems. Regardless of the distributed nature of the structure, the non-coordinated submodels can still share common features if there exist statistically consistent properties visible in the environment (see chapter 10).

There is no need of explicit coordination whatsoever — the mixture model is a simple extension of the linear case that was already shown to self-regulate and self-organize. As interactions with the environment are crystallized in the activity patterns, it is the feedbacks through the environment that assumedly again can accomplish the regulation task. What is more, all agents agree upon the goodness criterion — maximum activity and exploitation of the environment — and after that explicit coordination is no more necessary as the structure assumedly emerges from the competition. Whether or not some structure truly emerges in such a system is a difficult question — yet, the practical experience seems to support this hypothesis.

As a more abstract example, think of a formless social or memetic environment where it is difficult to uniquely quantify the structure or the variables. As studied in chapter 4, the neocybernetic view offers an escape here: It is the subjective individuals or individual minds that anchor the environment in the realm of observables. The population of minds determines the outlook of the constructivistic world, or the model for it — and, indeed, without this model the world itself would not exist!

The seemingly inaccurate and non-optimal mechanisms of representing the properties of individuals — genes in a biological system, and memes in a memetic one — seems to be nature's way to assure that not all submodels can end in the same local optimum. When there is no continuity among representations, separate individuals more probably produce different outcomes, ending in separate local minima of the cost criterion. Differences in genes span new directions in the high-dimensional property space, mutations perhaps augmenting this space, introducing new functionalities. Yet, there is some continuity, as it is the combination of the parent's genes that characterizes the offspring, mak-

[2]The ecosystem consists directly of the individual animals — the level of "species" is motivated for pragmatic reasons, because the spread of genetic information is limited by the species boundaries

ing the mechanisms of property inheritance more continuous, facilitating some level of simple parameter tuning even within a fixed structural framework. The genotype just determines the framework, and it is the dynamic interplay among the system and its environment that determines the outlook of the final phenotype. On the other hand, as the personal catastrophes (deaths of individuals) are non-synchronized, statistical properties of the population do not change abruptly, and the adaptation of the population becomes smoother. The nonlinear environment becomes modeled by local attractors; in a converged model the submodels are rather densely located, exhausting the information available in the environment more or less continuously.

The Darwinian mechanisms that come here to play to exploit the submodels, implementing the adaptation of the population level mixture model, good solutions among submodels being promoted in the mixture. The basic structure of an individual is determined by the genes, and within that framework, the familiar neocybernetic adaptation processes assumedly have tuned the parameters so that it is the best that can be achieved within that framework, so that the structures, and thus the underlying gene combinations, can be compared in an objective way. However, the idea of "survival of the fittest" is not so categorical as it is normally thought to be: Best solutions dominate, yes, but the outperformed ones also can survive, making the view of the reality more complete. Indeed, samples far from the mainstream solutions can carry very much valuable information. There are no outliers among the reproducing individuals, all models are valid: If an individual has survived so long, there must be something special about it; it is the whole adolescence that is there to filter out the actual mistakes. What is more, one needs to remember that the environment is not a predetermined entity, but it consists of other ever-adapting subsystems, and a stubborn individual can change its environment to make a new personal niche exist.

The role of birth and death are very central in Darwinian evolution theory. Now the system is more important than any individual; life is in the system, and in the population of individuals. As long as the system survives, there is no actual death. Another point is that because the genes only offer the pool of alternatives, the properties of an organism being mainly determined by the environmental conditions, one specific gene combination does not have such a crucial role.

Comparing to the Darwinian theory, again there is the fit criterion that plays a central role. However, now it is not about the search for the absolutely best fit — the population-level system searches for a *set of good fits* to implement a good mixture model, to better capture all aspects of the nonlinear environment. Indeed, the essence of modeling of the environment is not to find the actual winner, but to *find the definition of what fitness is* and map the whole "fitness landscape". And the primary reason for diversity is not to be prepared for the unknown future — the reason is simply to exploit the prevailing environment as efficiently as possible, now and here, with no future prospects. The traditional Darwinian thinking suffers from a intellectual discrepancy: Whereas the evolution mechanisms and fitnesses are defined on the level of individuals, the results are visible and meaningful only on the emergent level of the whole population. Whereas the lower and higher levels are traditionally incompatible, now both levels are combined in the same model framework, the individuals being submodels that together constitute the systemic model of the species — and the individual species further being submodels that together constitute the systemic model of the ecosystem. Thus, one can proceed from the analysis of individuals to analysis of populations, and from the analysis of species to the analysis of ecosystems; and if one can extend from the analysis of the existing taxonomies to the spectrum of possible ones, from characterizing details to seeing larger patterns, perhaps biology (and ecology) someday become real sciences.

There also exist less concrete populations where the same cybernetic ideas still apply. In a scientific world, for example, being capable of seeing similarities among individual paradigms and combining them in a larger model is similarly a central goal; rather than going deeper into the paradigmatic system, one tries to find more general systems connecting paradigms. In some environments the submodels need not be co-existing and parallel: The "populations" can be, for example, sequential, as it is often the case when speaking of human cultures. However, memetic systems leave signs of themselves, scriptures and artifacts, and as long as these signs can still be deciphered, faiths of various cultures can be reconstructed, and these cultures can be understood as consistent systems. Indeed, being based on such submodels, the highest-level memetic system can become alive — being manifested in a truly cultivated person. The human capacity blooms when one can put things in a perspective, constructing a balanced model of all aspects and dimensions of human culture: The human endeavor is to truly know what it is to be a human, and to understand how the human is connected to the world around him/her. It is not about memorizing details; it is about having a compact model where the individual facts have been combined into more general dependency structures.

6.4 More about sparse coding

For a moment, return to the linear case — it turns out that closer analysis gives insight to understand the general case, too, and the linear submodels efficiently support the emergence of the localization in the nonlinear global model.

It has been observed before that a cybernetic system implements principal component analysis, the submodels representing the (local) observations in terms of (global) variation structures. This is a simple result, as PCA is a mathematically rather trivial operation. Is there nothing else to be said about cybernetic data processing? Indeed, the PCA view is not the whole truth, it only determines the framework for data compression. Within the compressed data space, it is the selection of the latent basis that plays a major role when interpreting the results.

6.4.1 "Black noise"

In chapter 3, connections among $\bar{x}$ and $\bar{u}$, and among $\bar{x}$ and Δu were studied. When studying the theoretical mapping between $\bar{x}$ and the original undisturbed input u, it turns out that the eigenvalues of $\mathrm{E}\{\bar{x}\bar{x}^T\}$ can be expressed in terms of the n most significant eigenvalues λ_j of the original data covariance matrix

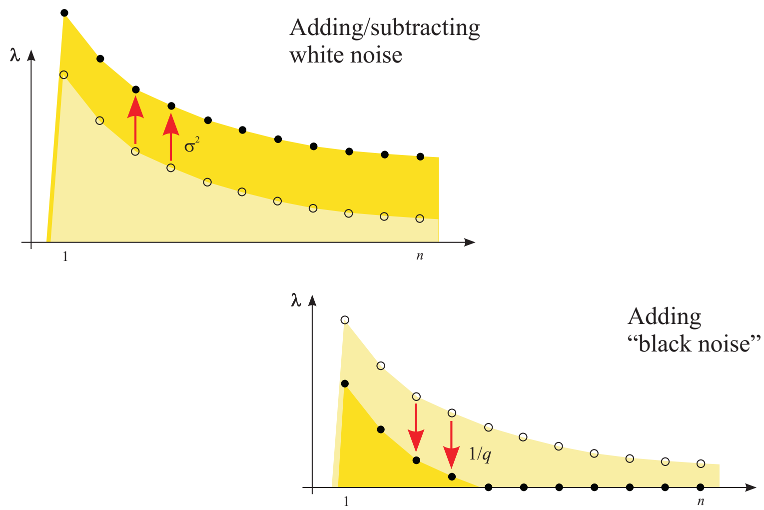


Figure 6.7: Consequences of adding "black noise" are opposite to white noise: The variation decreases in all directions — if possible

$\mathrm{E}\{uu^T\}$, as observed in chapter 3. Specially, if the coupling coefficients q_i and b_i are different for different neurons, the i'th eigenvalue (or latent variable variance) becomes

$$\frac{\sqrt{q_i \lambda_j} - 1}{b_i}, \tag{6.14}$$

indices i and j being ordered randomly. This reveals that there must hold $q_i \lambda_j > 1$ for that input variance direction to remain manifested in the system activity — if this does not hold, variable $\bar{x}_i$ fades away. On the other hand, for the modes fulfilling the constraint, interesting modification of the variance structure takes place; this can best be explained by studying a special case. Assume that one has selected $q_i = \lambda_j$ and $b_i = 1$ for all pairs of i and j. Then the corresponding variances become

$$\lambda_j - 1. \tag{6.15}$$

In each direction in the data space, the effect of the system is to bring the variance level down by a constant factor if it is possible (see Fig. 6.7). Analogically, because white noise increases variation equally in all directions, one could in this opposite case speak of "black noise".

What are the effects of this addition of black noise in the signals? First, it is the principal subspace of u that is spanned by the vectors ϕ_i. But assuming that this subspace is n dimensional, there exist many ways how the basis vectors can be selected, and some of the selections can be physically better motivated. For example, in *factor analysis* the PCA basis vectors are *rotated* to make them aligned with the underlying features, and the same idea takes place in *independent component analysis.* In factor analysis, it can be assumed that the underlying features can be characterized in mathematical terms applying the idea of *sparseness:* When a data vector is decomposed, some of the latent

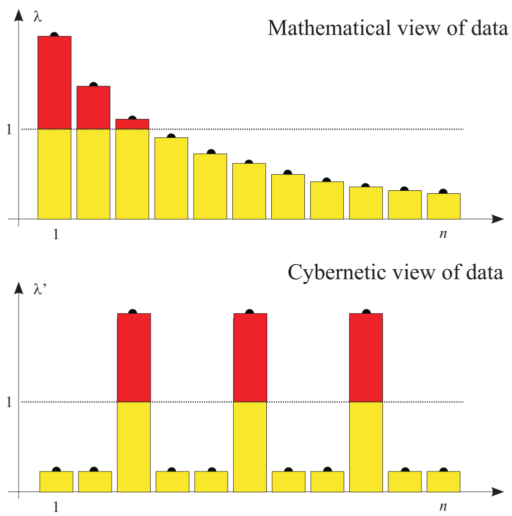


Figure 6.8: How black noise results in sparsity pursuit: Area above the threshold is maximized

variables have high scores while the others have low scores, increasing the differences among latent variable variances. This goal can be iteratively implemented in terms of criteria like *varimax* or *quartimax,* etc. In its extreme form, sparsity means that there are only a few of the candidates employed at a time, and the goal of modeling, rather than being minimization of the number of overall model size, it is the minimization of *simultaneously active constructs.* This means that the total dimension of the latent basis n can even become higher than the dimension m of the input data, the basis being *overcomplete.*

As shown in Figure 6.8, the *Hebbian feedback learning offers an efficient approach to achieving sparsity-oriented basis representation of the PCA subspace.* Whereas the overall captured variation (shown both in yellow and red color in the figure) is not changed by orthogonal rotations, the variation over the bias level (shown in red) *can* be changed. As the nominal PCA approach typically distributes variation more or less evenly along each latent variable, it is most of the variation that remains below the threshold level; now, as it is the area above the threshold level that is maximized, non-trivial basis representations are reached. When doing sparse coding, one can have $n > m$.

There are no closed-form expressions for implementing sparse coding for given data — there are only iterative algorithms. It seems that the algorithm proposed by the Hebbian feedback learning offers a compact and efficient alternative (see Fig. 6.9; compare to the algorithm in 6.2).

In the algorithm, the fixed states are first solved; because of the assumed linearity, infinite iteration changes into a matrix inverse. Actually, the linearity assumption here does not exactly hold: To make the sparse components differentiate, the cut nonlinearity is applied for $\bar{x}$, and, in principle, the matrix inversion does not give the fixed point (however, the system tends towards linearity; see below). The determination of `Xbar` is an extension of that in Algorithm 1, making the matrix inverse better invertible:

$$\bar{x} = \mathrm{E}\{I + \bar{x}\bar{x}^T\}^{-1} Q \mathrm{E}\{\bar{x}u^T\}\, u. \tag{6.16}$$

```
while ITERATE

  % Balance of latent variables
  Xbar = U * (inv(eye(n)+Exx)*Q*Exu)';

  % Enhance model convergence by nonlinearity
  Xbar = Xbar.*(Xbar>0);

  % Balance of the environmental signals
  Ubar = U - Xbar*Exu;

  % Model adaptation
  Exu = lambda*Exu + (1-lambda)*Xbar'*Ubar/k;
  Exx = lambda*Exx + (1-lambda)*Xbar'*Xbar/k;

  % Maintaining system activity
  Q = Q * diag(exp(P*(Vref-diag(Exx))));

end
```

Figure 6.9: **Algorithm 2**: Feedback Hebbian SCA by "selfish agents"

This is solved observing the loop structure, and exploiting (3.36). One can add the triangularization of the covariance matrix `Exx` here, too, to separate the components. The matrix Q is diagonal; "proportional control" with `P` as the control parameter is applied for (logarithms of) variable variations to keep the variation level of the variables in reference (`Vref` is the vector of reference values). Because the elements at the diagonal of Q are distinct, the components become distinguished, as discussed in chapter 3, and rather than implementing sparse subspace analysis, the algorithm implements sparse component analysis. Finally, after convergence the mapping of the model can be expressed as $\phi^T = Q\mathrm{E}\{\bar{x}\bar{u}^T\}$.

The neocybernetic algorithms can also be characterized in terms of mathematically compact formulas and theoretically powerful concepts. The sparse components represent (linear) submodels that together characterize a complex domain, perfectly matching the nonlinear case in 6.3.2. Summarizing, one can say conclude:

> It is the "clever agents" applying Hebbian/anti-Hebbian learning that implement theoretically correct principal component analysis that can be explicitly employed for theoretically optimal least-squares regression; the "selfish agents" applying feedback Hebbian learning implement sparse component analysis and simultaneously implicitly carry out robust regularized least-squares regression to control the environment.

This far all has been linear, the sparsity pursuit being implemented only through basis rotations. When the cut nonlinearity is included in the algorithm, cutting the minor (negative) variations explicitly to zero, only then the algorithm becomes strictly nonlinear. It turns out that the convergence properties of the algorithm can be enhanced considerably then. Because of the optimized rotations, one already has minimized the cross-cluster effects, and for "typical" data located in such clusters, there probably are no crossing-overs between linear submodels. Structure changes are located in deserted regions in space, and rather than being piecewise linear, the model is "practically linear". In the converged system, the role of nonlinearity is rather transparent. But there is more.

The nonlinearity that is introduced in the structure does *not* make the system essentially more complicated. When studying closer the data processing (again see Fig. 3.3), it is interesting to note that the nonlinearity that is now applied is *outside* the inner loop, just filtering the incoming information. The basic functionality of the system is still determined by the closed loop as shown in the figure, converging so that the best possible linear matching between the realized $\bar{x}$ and Δu is implemented, however these signals are externally deformed. This means that despite the nonlinearity, the model tends back towards linearity and statistical optimality.

6.4.2 Towards cognitive functionalities

Modeling of the environment is common to all cybernetic systems. The properties of the environment — like nonlinearities — are best quantifiable when the system resides in infosphere, the signals being better commeasurable, and the existing data structures are intuitively comprehensible.

Example: Modeling of biped walking

When studying the geometric structure of limbs, it is evident that the dynamic model for them is highly nonlinear. Still, to keep a two-legged body stable, very precise control is needed. Whether such control structures can be based on linear submodels that are tuned applying measurement data, was studied in [34]. The available data consisted of state vectors characterizing the orientation and velocities of a simplified two-legged structure and its relation to the surrounding world. The nonlinearities in the adopted model structure were distributed in substructures; it was assumed that the nonlinearities are smooth, and "nearby" data samples share the same locally linear model — that is, the observation data was first clustered, and data within each cluster was used to construct a local linear model. Because of the high assumedly redundant dimensionality of the data, the linear models were based on PCA compression of the observation data, and the motion controls to achieve the walking gait were thereafter reconstructed applying principal component regression based on that model.

It turned out that the clustered model could reproduce the motion controls in a satisfactory manner, and the simulated motion remained in control. However, the model was not quite satisfactory: From the cognitive point of view, the model structure was not very plausible. There was the predetermined structure with separate levels of inter-cluster and intra-cluster operation — coarse

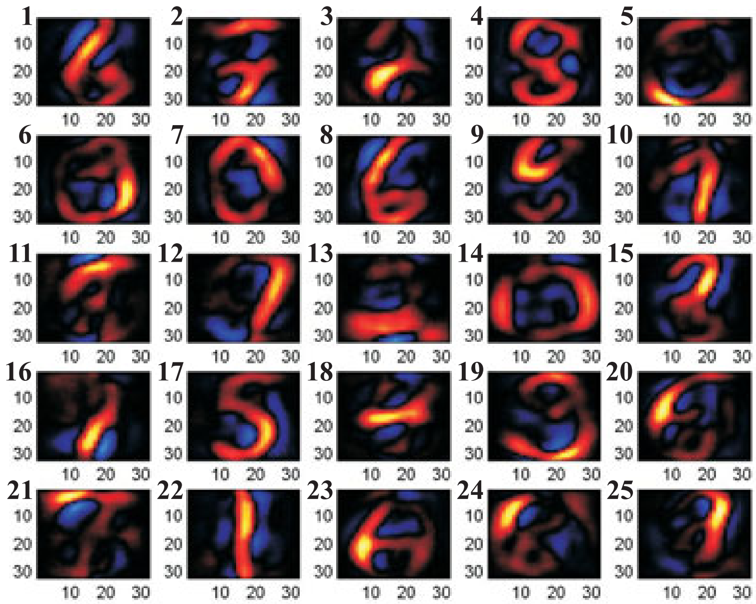


Figure 6.10: The 25 sparse components extracted from the handwritten digits (random ordering). It seems that different kinds of "strokes" become manifested (see below)

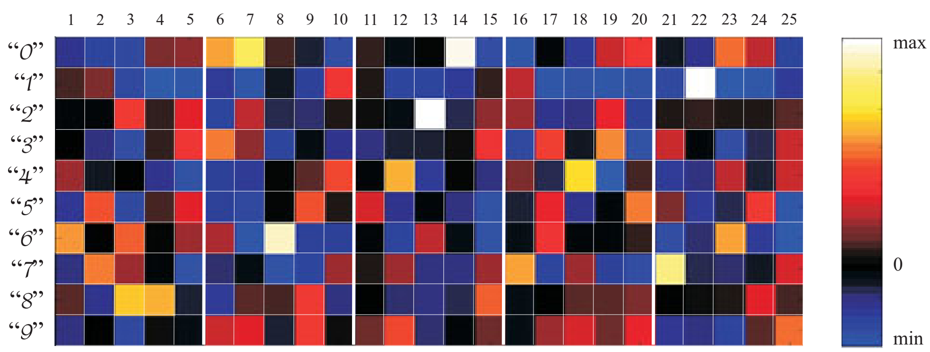


Figure 6.11: How the different digits (left) are represented by the 25 sparse-coded features (above). For example, it seems that feature #14 is only active when the input pattern is "0", feature #22 when it is "1", etc. Feature #13 correlates strongly with "2", as the other patterns seldom occupy the bottom rows. Some patterns have various alternative forms (like "3" is represented either by the feature #6 or feature #19). For most of the input patterns there are no unique matches — they must be composed of parts (for example, "4" seems to be a sum of features #12 and #18). Whether or not the features are disjunctive or conjunctive is determined by the optimization machinery as the data is processed

matching against the clusters, and after that the fine tuning against the cluster-specific submodels. It seems that some higher-level control is necessary here during the model usage as well as during model adaptation. However, it turns out that this is not the case.

As discussed in chapter 7, a grid of linear Hebbian neurons implements the neocybernetic model, modeling the nonlinear environment. Hebbian feedback learning implements the PCA compression of the data, constructing a sparsity-oriented model. Sparse coding results in differentiation of the substructures, or emergence of localized "clusters". Within this framework explicit control of clustering or selection among submodels can be avoided because of the competition among substructures, the best matching submodel automatically receiving most of the activation. There is contribution also by the lesser submodels — this means that there is smooth transfer between submodels in the data space.

What comes to the cluster-based representation of nonlinearities, there is also no need for additional functionalities in the neocybernetic framework. Still, there are challenges: How to implement the input–output structure so that the regression onto the control signals can be implemented in a plausible way? And how to implement optimization towards smoother and faster movements beyond the available prior behaviors?

Structures in infosphere

To illustrate the structure based on sparse codes in more abstract terms, again study the case of hand-written digits (see Sec. 6.2.2). Each of the latent variables $\bar{x}_i$ was kept active by appropriately controlling the coupling factors q_i. Figure 6.10 shows the results when applying the presented algorithm (see also discussion in Fig. 6.3), and Fig. 6.11 presents how the converged features were oriented towards the input patterns. Note that the goal of this coding is not to distinguish but to find similarities — that is why the received feature model is probably not good for classification tasks.

The behaviors in this experiment differed very much from those when applying principal component coding: During the convergence process, in the beginning, something like clustering emerged, each data region being represented by a separate variable; as adaptation proceeded, the features started becoming more orthogonal, and patterns were decomposed further. What is interesting is that this kind of "stroke coding" has been observed also in the visual V1 cortex region in the primate brain (see [29] and [59]): It seems that the visual view is decomposed into simpler, statistically relevant substructures.

What if more complex data is modeled applying the same kind of sparse coding schemes? This was studied using *textual documents.* There were some hundred short descriptions of scientific reports on different aspects of *data mining.* Very simple representation of the texts was selected: It was assumed that the documents can be characterized by the set of words that is found in their descriptions. Data dimension was huge as there was one entry for each of the words in the data vectors. The document were represented by their "fingerprints", or data vectors containing their word counts. After some data preprocessing (see [92]), sparse coding was applied, and the resulting sparse structures representing the correlation structures among the words are shown in Fig. 6.12. It seems that

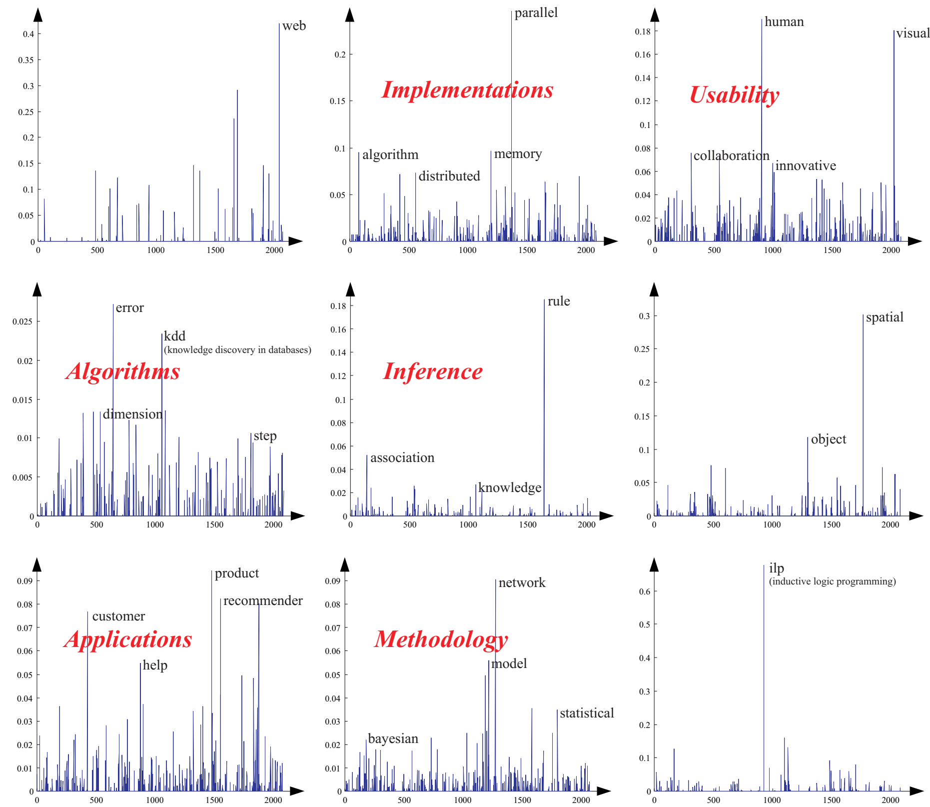


Figure 6.12: Results when textual material (documents on "data mining") are modeled applying sparse coding techniques: It seems that the emerging data structures capturing the correlation structures among the words are *generalized keywords* characterizing the different dimensions in the documents. The nine keywords are projected against the original words that are listed on the horizontal axis in alphabetical order; long bars denote high relevance. The keywords are named afterwards after studying the semantics of the words characterizing them.

the extracted data structures can be used to bring structure even to this semantically complex domain: Different documents can be represented as weighted combinations of the contextual "strokes".

Even though one should be careful about too strong conclusions, these experiments still motivate excursions to truly challenging domains of cybernetics, namely, to the world of cognitive systems — this is done in the next chapter.

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

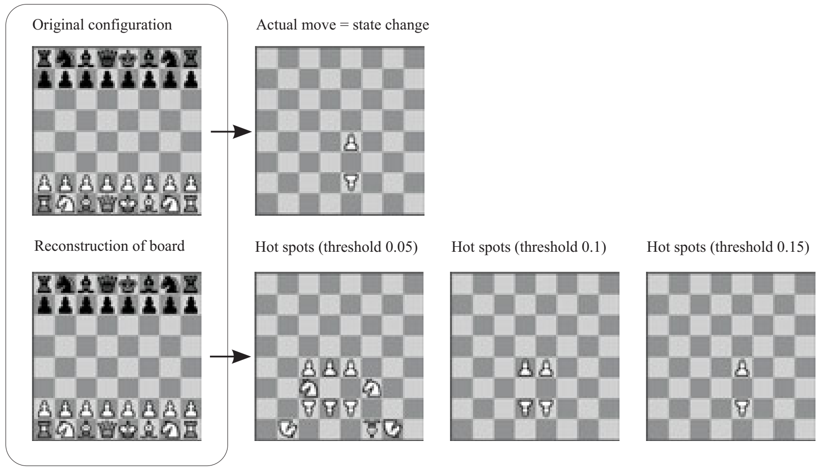


Figure 7.1: Reconstruction of the board and visions of the future in the beginning of the game

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

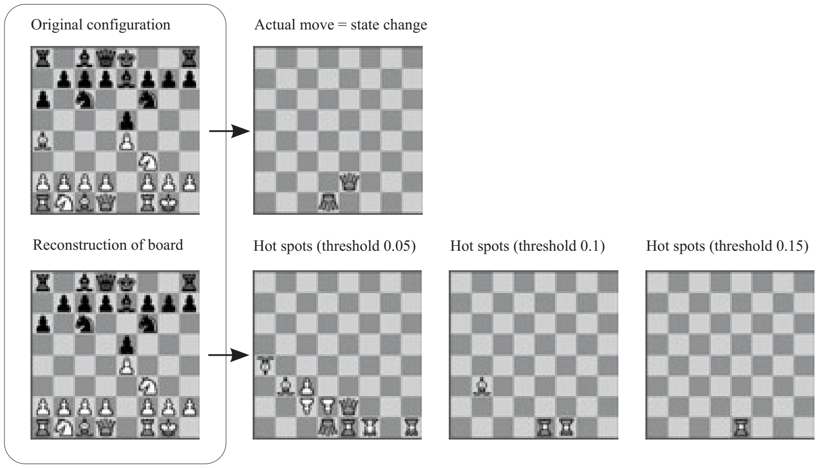


Figure 7.2: Reconstruction of the board and visions of the future at some later phase in the game

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"

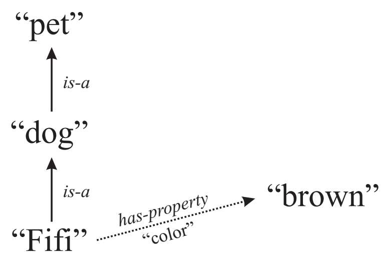


Figure 7.3: How different concepts determine each other: *Semantic net*

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

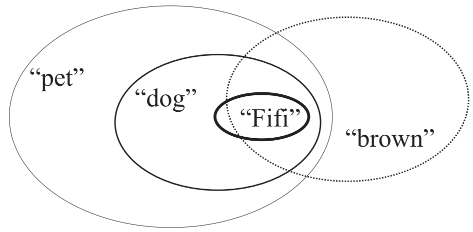


Figure 7.4: How different concepts determine each other: *Fuzzy subsets*

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

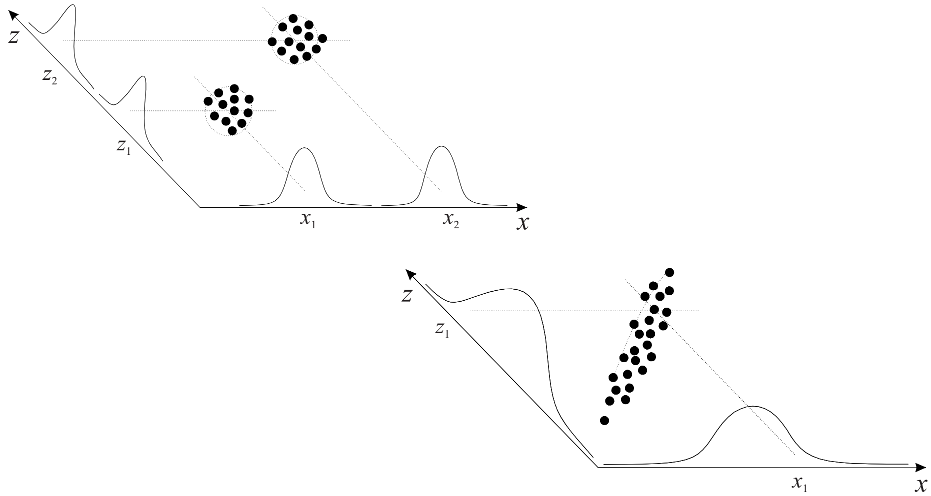


Figure 7.5: Traditional view of expertise (on the top left) makes it possible to implement rules of the form `IF` $x = x_i$ `THEN` $z = z_i$, etc., whereas when distributions are employed, inference becomes subtler, being an associative (maximum likelihood) pattern matching process against the existing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

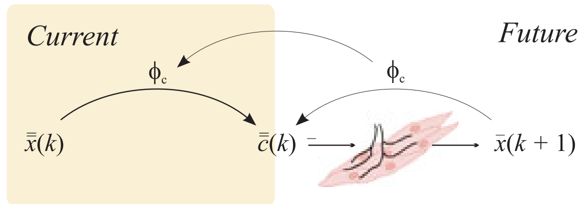


Figure 7.6: Control as a (antisymmetric) mirror between the past and the future

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 $\bar{\bar{c}}(k) = \phi_c \bar{\bar{x}}(k)$, 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 $k \rightarrow k+1$, 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

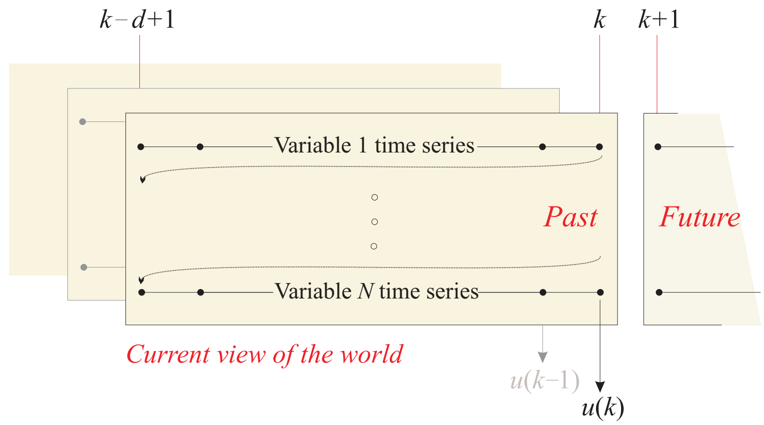


Figure 7.7: How the sensory registers can capture the state of the world

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

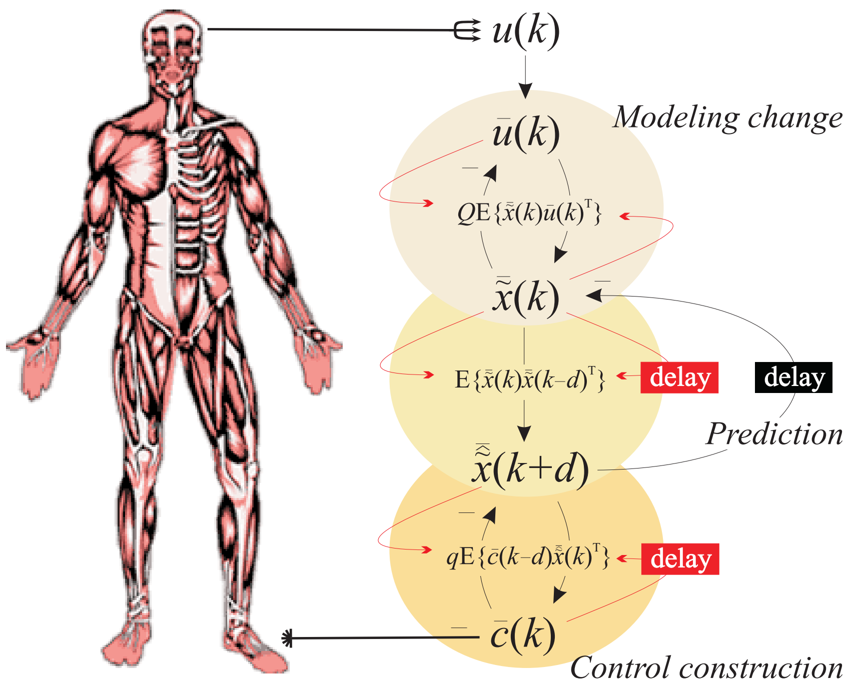


Figure 7.8: "Almost complete" solution to cybernetic control

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

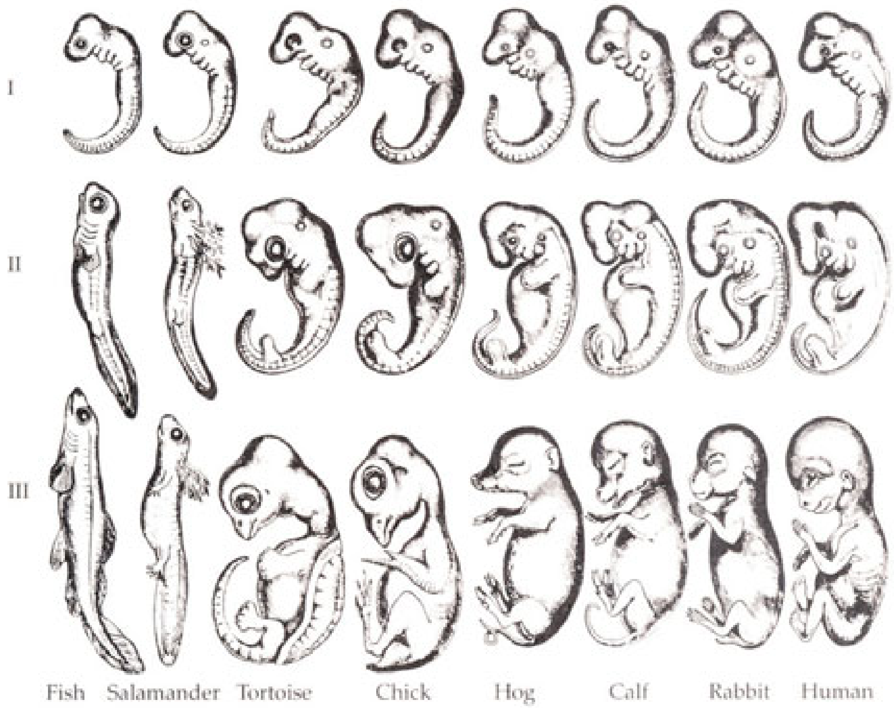


Figure 7.9: Graphs that are today usually called "Haeckel's lie". Ernst Haeckel claimed that the embryos of fish, chicken, human, etc., are evolving through the essentially same phases, repeating the common history of species development

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.

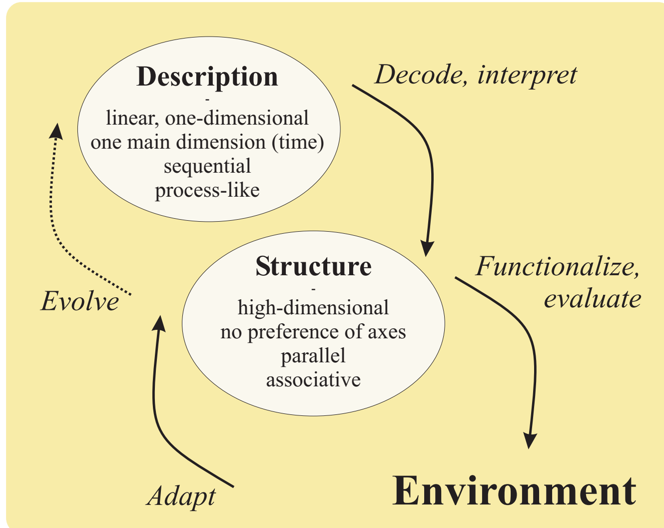


Figure 7.10: From sequential to parallel — each representation being meaningful and interpretable only in an appropriate environment

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?

Level 8

From Building Blocks to *Theories of Everything*

When studying the cognitive system in the neocybernetic perspective, as was done in the previous chapter, some interesting results can be found. However, when the semantic grounding is left floating, so that concepts are defined contextually only in terms of other ones, there is the annoying question haunting: "So what?" The resulting models are just computational constructs; intuitively they cannot have very much to do with "real" intelligence that has to be grounded on flesh and bones. This problem does not only apply to the cognitive systems: Generally, one has problems if trying to capture behaviors in complex domains where semantics is detached. The lower-level system one is studying, the more there is need to attack the deep coupling between the system and the real world. Only if the domain-area semantics is to some extent captured, the computing machinery can reveal something relevant and non-trivial that has not been programmed in the code in the beginning.

What are these unexpected results — and what does this capturing of semantics mean? The problem here is that the essence of domain-area semantics assumedly is different in different domains, and no generic approaches perhaps can be determined. However, specific examples can be illuminating, giving hints of what cybernetics is all about after all.

Here, a very detailed case is studied, the environment with its semantics being implemented; the environment is that of complex molecules, and the application domain is modeling the sequences of amino acid sequences corresponding to the translated genetic code. The sequences of atoms in the molecules are codes themselves, and the objective here is to interpret that "language of molecules" that is interpreted as functionalities in the chemical environment. From the point of view of understanding living systems, the case of amino acids is crucial, because they determine how the final functional proteins are folded, thus dictating their structure and function. The protein structures are the basis for all biochemical processes, and they are the basic structures in all living organisms. Understanding such codes is the key to real "bioinformatics".

And if the principles of life can be reduced back to quantum physics, perhaps there is life in large-scale physical systems, too? The wild speculations in the end reveal the value of analogues as a source of intuition.

8.1 Computationalism cybernetized

Ludwig Wittgenstein observed that the language determines the limits of one's thinking: If there are no appropriate concepts, there is no way to express oneself. Similarly, one cannot discuss complex systems without appropriate concepts. However, in the adopted evolutionary framework, this thinking goes much deeper: No humans are needed there to use the language; the language is there for interaction with the environment, it is the environment that reads and interprets the code. There is a code, based on a special language, to make it possible for the system to "express itself" — or, indeed, in very concrete terms, the system is defined in that language. Without domain-oriented languages appropriate constructs and interpretations cannot be defined, and the systems do not become alive in their environments.

8.1.1 Formal and less formal languages

The essence of evolution, or any developmental processes, can be represented as code — or, anyway, different kinds of codes is the way how nature does it. The brain is not a unique medium of decoding languages — perhaps it is the most versatile, but qualitatively the mental system is by no means alone in its aspirations towards capturing the complexity of evolutionary systems. The deep structures of the language [18] are different in different domains. What do we know about such representations?

In addition to the natural languages, there exists a wealth of formal languages being used today. It is instructive to study the special case of *programming languages,* as there seems to be evolution among them, too, from clumsy towards more natural ones. The larger the programs have become, the more structured the formalisms have become, being better capable of representing different substructures in the observed world. In modern programming languages, the computation has been "packaged" tight, the control of computation being distributed; the programming languages have evolved through procedural languages to the today's object-oriented ones. The development of programming languages seems to lead towards representations that more and more match the mental structures: In modern languages, there are "classes" that stand for categories, "objects" being individual instances or examples representing the class. The implementations of the attributes, or the "methods", however, are rather different. To make the programming languages more useful, different kinds of sequential control structures are employed; parallelity or fuzziness in data processing, however, is not heavily addressed, reflecting the dominance of centralized thinking. Of course, there are also the pragmatic reasons for the shortcomings, the processors in today's computers operating one instruction at a time in an all-or-nothing fashion.

The programming languages, as well as the other code systems, are used to describe the desired functioning of the world. In the general-purpose programming languages, all this functioning needs to be implemented explicitly in the code. In the other extreme, one can think of a code that only determines the structures, the functioning being supplied by the language interpreter: The environment can carry out the functionalizing, or "waking up", of the non-living structures, if it supplies ready-to-use hooks to existing functionalities, or dynamic attractors just waiting to be activated — that is, if the code matches with the semantics supported by the environment. Codes just select among the candidate attractors to put up the system.

By definition, however, formal languages are formal or syntactic, missing semantic content. Just as in formal logic, the syntax is separated from the details of the domain field to keep the structures general. In the other extreme, there are the spoken natural languages that are overfull of semantics, being loaded with non-formalizable nuances. A speech act consists of not only the actual utterance but there are, for example, the facial expressions and gestures, and there can be the spices of humor and irony accompanied. The (collective) mental system has developed natural language for communication face to face, not for losslessly storing and transmitting information in text form. Are there other kinds of "somewhat natural" languages that would be appropriate for representing the coordinated developmental processes in other phenospheres? Such a cybernetic language should be some kind of a compromise between formal and natural, capturing a narrow domain with restricted semantics, offering a window to the attractors of dynamic low-level processes in the environment.

To understand the challenges being faced, another aspect about truly natural languages needs to be pointed out. As studied in chapter 3, the key observation there was that everything is implemented locally by uncoordinated actors. This fact applies here, too: The exact-looking representations are not so exact, they are just the emergent nominal patterns. The same applies to codes — meaning that one needs to master the regions within and between the codes. In principle, all diversions from the exact code are errors, but, in practice, there are "degrees of impossibility". As there is normally just one way to interpret the codes, new innovations cannot be studied: Adding noise just breaks the structures, making codes completely undecipherable. One needs (and nature needs) a possibility for "domain-oriented noise", where alternative routes of evolution can be taken in reasonable directions to escape the current stasis. It has been observed that random mutations practically never result in enhancement of the genetic code — it is like adding typing errors in a book: If the text can still be read, there are no meaningful changes in the contents, there are no new memetic structures. Even though the code is represented as a linear list, its structure is hierarchic, and there is huge difference in the effects depending on whether the mutations are in the "leaves" of the hierarchy, or in the "root".

To have more understanding of languages that are not so fixed to the distinct symbols, and to make the topology among the constructs visible and analyzable, wider perspectives are necessary. One needs to implements also "metainformation" for reading the information in the actual code. Only looking at the language is not enough as this metainformation concerning the code does not exist in the code itself — it is in the interpreter mechanism, or in the environment.

To determine an appropriate interpretation of the code, the semantics of the environment has to accompany the language. Actually the key point is not the language but this interpreter, or compact representation of the application domain. How to determine the attractors that are relevant in the domain to be employed by the systems? If this can be done explicitly, computers can be employed, and the "seminatural" language changes to a programming language, facilitating analyses of complex system *in silico*. The computing power alone does not help if the computer does not do the correct things, but if the environment is implemented appropriately, computationally keeping the dynamics running, supplying the domain-specific dynamic attractors, *simulation* of the development processes can be carried out.

Wolfram [91] says that tomorrow's science is based on simulation. Even though his basic hypotheses is not true (cellular automata are not the only available model family), perhaps he is not totally wrong. Simulation is the way to escape the formal rules and formulas in the theories — but it is not simulation of the system itself, it is simulation of the environment that is needed. Traditional theories are still needed — first to implement the domain-area semantics, and after simulations for compiling the results. Simulation is the method of creating data, fresh processes supplying alternatives or local solutions, the final cybernetic model then being composed over the map of alternatives in the environment.

8.1.2 Simulators of evolution

Evolution is more that mere adaptation in an environment; it involves structural change taking place after the initiation of the system. If this kind of evolution takes place within a single individual it can be called development — in any case, modeling such processes is crucial when trying to capture the essence of biological systems. Unfortunately, there is no strong topology for systems where structures evolve. When seen in the mathematical perspective, the changing structures can be characterized as the processes being highly nonlinear; there are typically no closed-form mathematical solutions to them. A robust and generic approach "homogenizing" the details of nonlinearities is iterative: The models can always be simulated.

Just as the nonlinear representations, codes, too, can have very different outlooks — for example, there are many natural languages that still span the same mental view. There do not necessarily exist any one-to-one mappings between the surface forms of languages — their correspondences can only be evaluated in terms of semantics, through the deep structures, or the activated attractors in the appropriate environment. Only in this sense uniformity among representations can be reached: Codes are identical if they result in the same system of internal dynamics. General analysis of codes and their correspondences can only be carried out through the simulation of the environment — perhaps this is the way towards a general theory of "natural linguistics". To make the syntax and the semantics go together, the point of view needs to be changed. It is not the code that is simulated; it is the environment that runs, and when codes are put in, they are interpreted in that process, so that the operational systems emerge from the simulation. Simulator of the environment remains intact, the systems changing; the special challenge is caused by the fact that changing of the systems can also change the environment possibly introducing new attractors and modifying the old ones. Of course, the codes can be put in the potion only after the connection between the language constructs and the existing dynamic attractors are defined.

As is evident, the implementation of the simulators in different environments are different, but there are some general ideas. All complex environments share common characteristics: One of the challenges — nonlinearity — was discussed in chapter 6; The other challenge — time-varying nature — was studied in chapter 7, and also in this chapter. To support nonlinearity, one needs to support a population of individual processes, where each of such submodels takes care of its personal local attractor, or local minimum of the cost criterion; to support varying in time, the simulator has to decode the instructions changing the system structures in a coordinated way. To summarize, the simulator has to simultaneously (fractally!) host different types of processes:

- To tackle with the nonlinearity of the world, there have to be *parallel* processes.
- To tackle with the time-variability of the world, there have to be *sequential* processes.

The contribution of the cybernetic thinking here is to offer intuitions. For example, how the population of competing submodels can be maintained and kept stable — of course, this control is localized, coordination being implemented implicitly through the environmental feedbacks, and the neocybernetic frameworks offers an off-the-shelf framework. The other cybernetic contribution is the emphasis on balance pursuit: At each level, dynamic equilibria are searched for — it is these balances that are the dynamic attractors, or the domain-area "concepts". For example, in the case below, the domain-area semantics is implemented based on the neocybernetic model. One can concentrate on such important issues rather than details in simulations.

In short, the role of the simulation machinery is to host the proto-systems in an appropriate environment, shuffle this container, and see the spectrum of outcomes — just as one would do with *in vivo* experiments. It is the computer that now supplies for the mindless signal carriers that operate on the structural building blocks to construct the final systems. The population of surviving systems characterize the local solutions, representing the distribution to be analyzed. The process of "running the programs" in such an environment is a form of *Markov Chain Monte Carlo* simulation. As compared to today's approaches for doing this kind of first-principles simulation, there are only minor differences: For example, the models based on the basic theories of quantum theory are too involved to deliver enough relevant information as the simulations are very computationally demanding; now one does not need such detailed calculations, as it is not only shuffling and cumulating of numeric errors. When one concentrates on the relevant attractors only, one has convergent rather than divergent models, and errors due to simplifications fade away. But what is this "emergent-level quantum physics"?

All this sounds magnificent — but can such mechanization of semantics ever be carried out in any interesting domain? In what follows, the basics of physical chemistry are studied. Study what kind of consequences it has if a molecule is regarded as a (truly) cybernetic population of charge fields; try to constitute an absolute grounding of semantics in such domain, define appropriate data structures that seamlessly reflect the underlying realm, and define the rules for combining the data structures. The pursuit for semantics must be extended to the very kernel — indeed, when studying physico-chemical systems, it is *quantum theory* that is addressed. From the point of view of understanding biological systems, this problem setting is very relevant: For example, if the code is the sequence of amino acids, its interpretation is the three-dimensional, folded and functioning protein molecule — and this is the nature's way of implementing structures that are defined in the genes. In this environment, the translation of a sequential code into a high-dimensional structure is implemented in very concrete terms. This has to be seen as a (very) preliminary sketch of what the actual implementation of a "proteomic simulator" could look like.

8.2 Emergence in a physical system

Erwin Schrödinger, one of the pioneers in modern physics expected and hoped to find new physics through a study of life [69]. He saw a deep connection between the high-level systems and the low-level principles. Also since him, similar more or less well-grounded intuitions have been exploited: For example, the mysteries of cybernetic processes — not only life itself, but cognition, etc. — have been reduced into the twilight of elementary particles, hoping that "free will" emerges from the unpredictability of quanta. Perhaps there exist more concrete contributions, too — fundamental physics is after all the realm where the living chemical systems reside, and it is here where the system semantics, and "grammar" of the codes, has to be based on.

8.2.1 Cybernetic view of electrons

There is no central control among individual electrons, but the electron systems — atoms, molecules — still seem to be stable and organized. Either there is some yet unknown mechanism that is capable of maintaining the stability and the structures — or, it is the neocybernetic model that applies. The latter assumption is now applied, and the consequences are studied. The starting point (set of electrons) and the goal (cybernetic model) are given, and the steps in between need to be motivated; of course, this is a risky way to proceed, as everything is interpreted in a predetermined way. Results are not conclusive, the goal is just to present an idea and an approach that is offered by the adopted neocybernetic framework as there can be new useful intuitions and interpretations available. Perhaps these studies can be motivated using the words of Max Born:

> All great discoveries in experimental physics have been made due to the intuition of men who made free use of models which for them were not products of the imagination but representations of real things.

So, assume that the nuclei are fixed (according to the Born-Oppenheimer approximation), drop the electrons in the system to freely search their places, and see what happens.

When studying the elementary particles, traditional thinking has to be turned upside down: For example, it seems that in that scale the discrete becomes continuous, and the continuous becomes discrete. Distinct electrons have to be seen as delocalized, continuous charge distributions; however, their interactions have to be seen not as continuous but discrete, being based on stochastic photons being transmitted by the interacting charge fields. This view needs to be functionalized.

Assume that there are two (non-measurable) charge fields i and j, variables x_i and u_j representing their momentary intensities. These fields are manifested through the photons emitted by them; the probability for a photon to be transmitted is proportional to the field intensity. For two fields to interact, the photons need to meet — assuming that the photon transmission processes are independent, this interaction probability is proportional to the product of the intensities, or $x_i u_j$. However, such momentary elementary interactions cannot be detected; the macroscopic phenomena are emergent and become analyzable only through statistical considerations. It is *electric potential* that is such an emergent phenomenon, assumedly being a longer-term average of interactions over time.

To estimate the energy that is stored in the potential fields, one can calculate $p_{ij}\, x_i u_j$, where p_{ij} is the overall probability of the two fields to overlap. Because the field has a dual interpretation, also representing the probability for a charge to be located there, one can estimate the probability of coexistence as $p_{ij} = \mathrm{E}\{\bar{x}_i \bar{x}_j\}$ when the two charge fields are assumed independent. Energy is a scalar quantity; when there are various overlapping charges, their total potential can be expressed as a sum $\sum_{i,j} p_{ij} x_i u_j$, or when expressed in the matrix form, $x^T \mathrm{E}\{\bar{x}_i \bar{x}_j\} u$. However, there are different kinds of charge fields, attractive and repulsive. Assuming that the vector x contains the negative fields, representing the electrons, and u represents the positive charges, one can write for the total energy

$$J = \frac{1}{2} x^T \mathrm{E}\left\{\bar{x}\bar{x}^T\right\} x - x^T \mathrm{E}\left\{\bar{x}\bar{u}^T\right\} u. \tag{8.1}$$

In the former term there are the self-referential structures (there is potential energy stored also when a single charge field is put together), and its outlook can be motivated as in 3.1.2. The key point here is that when appropriate interpretations are employed, it is the neocybernetic cost criterion that is found, meaning that the solutions for electron configurations also implement the same neocybernetic structures. If the assumptions hold, there is self-regulation and self-organization among the electrons, emerging through local attempts to reach potential minimum. Not all electrons can go to the lowest energy levels, and "electronic diversity" emerges automatically. Surprisingly, because of their delocalization, "overall presence" and mutual repulsion, the electron fields implement explicit feedback, following the model of "smart cybernetic agents" (see 5.2.2).

It is interesting to note that there are no kinetic energies involved in the energy criterion, and no velocities or accelerations are involved. As seen from the system perspective, the charges are just static "clouds". This means that some theoretical problems are now avoided: As there are no accelerating charges, there are no electrodynamic issues to be explained as no energy needs to be emitted, and the system can be in equilibrium. In contrast, such electrodynamic inconsistencies plagued the traditional atom models where it was assumed that the electrons revolved around the nucleus, experiencing constant centripetal acceleration, so that radiation of energy should take place.

Whereas the electrons are delocalized, the heavier nuclei can be assumed to be better localized. The key observation here is that the analysis of the continuous space — modeling of the charge distribution of electrons — changes into an analysis of a discrete, finite set of variables, or the nuclei. The idea of "mirror images" is essentially employed here — rather than studying the system itself, the electrons, its environment is analyzed: In this special case it is the environment that happens to be simpler to operate on. Assuming that the interactions among the distinct nuclei can be represented in terms of a covariance matrix $\mathrm{E}\{\bar{u}\bar{u}^T\}$, the charge distributions of electrons are revealed by its eigenstructure. When the eigenvectors are denoted as ϕ_i, one can rename the constructs: The "orbits" of electrons, determined by the eigenvectors, are discretized (molecular) orbitals. Following the Max Born's formalism, the "squares" of orbitals represent "real" probabilities, or actual charges — and, indeed, the basic quantum theoretical assumptions seem to hold: For example, the integrals, when changed to summations, again equal

$$\sum_{j=1}^{n} |\phi_{ij}|^2 = \phi_i^T \phi_i = 1. \tag{8.2}$$

Because of the properties of eigenvectors, the discrete orbitals are mutually orthogonal. Traditionally, it is assumed that there is just room for a unique electron in one orbit (or, indeed, for a pair of electrons with opposite spins). However, now there can be many electrons in the same orbital, and there is no need to employ external constraints about the structures, like assumptions of spins, etc. The charge field can be expressed as $\psi_i = \sqrt{\lambda_i}\,\phi_i$, so that the overall charge becomes $\psi_i^T \psi_i = \lambda_i$. The "variance" λ_i is the emergent measurable total charge in that field. This means that there are some conditions for the charge fields to match with the assumption of existence of distinct charge packets:

1. The eigenvalue λ_i has to be an integer times the elementary charge, this integer representing the number of electrons in that orbital.

2. The sum of all these integers has to equal the number of valence electrons, sum of all free electrons in the system.

These constraints give tools to determine the balance configuration among the nuclei (see later). As studied in 3.2.3, the energy drop in the orbital is related to λ_i^2.

How to quantize the continuous fields, and how to characterize the effects in the form $\mathrm{E}\{\bar{u}\bar{u}^T\}$, and how to determine the parameters? And how is this all related to established quantum theory? In short, how are the above discussions related to real physical systems? These issues are studied next.

8.2.2 Molecular orbitals

Atoms are already enough well understood — at least what comes to the hydrogen atom (!). The contemporary theory of atom orbitals can explain their properties to sufficient degree. However, it seems that one needs new approaches to understand the emergent level — or the level of molecules. Molecular orbitals are interesting because the chemical properties of compounds are determined by their charge distribution — essentially these orbitals reveal how the molecule is seen by the outside world.

The molecules have been a challenge for modern physics for a long time, and different kinds of frameworks have been proposed to tackle with them: First, there are the *valence bond theories,* where the individual atoms with their orbitals are seen as a construction kit for building up the molecules, molecule orbitals being just combinations of atom orbitals; later, different kinds of more ambitious *molecule orbital theories* have been proposed to explain the emergent properties of molecules. In both cases it is still the ideas of atom orbitals that have been extended to the molecules. Unfortunately it seems that very often some extra tricks are needed: for example, to explain the four identical bonds that carbon can have, peculiar "hybridizations" need to be employed; and still there are problems, a notorious example being *benzene* (and other aromatic compounds) where the "bottom up" combinations of atom orbitals simply seem to fail. And, unluckily, it is exactly carbon and its properties that one has to tackle with when trying to explain living systems and their building blocks.

When thinking of alternative approaches, it is encouraging that molecules have been studied applying discretized eigenvalues and eigenvectors before: For example, Erich Hückel proposed an approach that is known as *Huckel's method,* also reducing the analysis of energy levels in molecules into essentially an eigenvalue problem [8]. However, this method is still based on combinations of atom orbitals, and being based on crude simplifications, it is regarded as an approximation. It is also quite commonplace that linear additivity of orbitals is assumed on the molecular level — normally it is atomic orbitals that are added together, now it is molecular orbitals directly. Indeed, basic physics *is* linear; the problems are normally caused by the huge dimensionality of the problems.

Now it is assumed that all of the molecular orbitals extend over the whole molecule, and it is assumed that (8.1) characterizes the electrons; the challenge is to combine this with current theories and models. It is the *time-independent Schrödinger equation* that offers a solid basis for all quantum-level analyses [10]. It can be assumed to always hold, and it applies also to molecules (h is the Planck's constant, and m_{e} is the mass of an electron):

$$-\frac{h^2}{8\pi^2 m_{\mathrm{e}}}\frac{d^2}{dx^2}\psi(x) + V(x)\psi(x) = E\psi(x). \tag{8.3}$$

Here, $V(x)$ is the potential energy, and E is the *energy eigenvalue* corresponding to the *eigenfunction* $\psi(x)$ characterizing the orbital. As $\psi(x)$ is continuous, Schrödinger equation defines an infinite-dimensional problem, and as x is the spatial coordinate, in higher dimensions this becomes a partial differential equation. Normally this expression is far too complex to be solved explicitly, and different kinds of simplifications are needed. Traditional methods are based on reductionistically studying the complex system one part at a time, resulting in approaches based on the atom orbitals. Now, start from the top: As studied in the previous section, assume it is simply a non-controlled play among identical electrons that is taking place in a molecule. It is all "free" electrons that are on the outermost shell that are available for contributing in the orbitals, that is, for a carbon atom the number of valence electrons is increased by the number $v_{\rm C} = 4$, for hydrogen $v_{\rm H} = 1$, and for oxygen $v_{\rm O} = 6$ (!). What kind of simplifications to (8.3) are motivated?

The time-independent *discrete* Schrödinger equation that is effectively being studied is defined now as

$$-V_0\phi_i + V\phi_i = E_i\phi_i, \tag{8.4}$$

where ϕ_i are now vectors, $1 \leq i \leq n$, dimensions equaling the number of atoms in the molecule n; because of the structure of the expression, these are the eigenvectors of the matrix $V - V_0$ corresponding to the eigenvalues E_i. Comparing to the discussions in the previous section, there holds $E_i = \lambda_i^2$, the eigenvectors being the same. Rather than analyzing the infinite dimensional distribution of electrons study the finite-dimensional distribution of nuclei; one only needs to determine the $x \times n$ elements of the potential matrix $V - V_0$ to be able to calculate the orbitals (or the negative charge fields around the positive nuclei).

To determine the matrix of potential energies among the nuclei, the challenge is to determine the terms corresponding to the first term in (8.3). The diagonal entries of $V - V_0$ are easy: Because the "local potential" is assumedly not essentially affected by the other nuclei, the atoms can be thought to be driven completely apart, so that the non-diagonal entries vanish; the diagonal entries then represent free separate atoms, so that the electron count must equal the number of available valence electrons, that is, the i'th diagonal entry is proportional to v_i^2, where v_i presents the number of valence electrons in that atom. For non-diagonal entries, the sensitivity to changes to distant nuclei becomes small, so that the term with the second derivative practically vanishes, and the corresponding entry in the potential energy matrix is according to basic electrostatics approximately proportional to $\frac{v_i v_j}{|r_{ij}|}$ without normalization. Here, $|r_{ij}|$ stands for the distance between the nuclei i and j. When the preliminary potential matrix has been constructed, elements of the matrix $V - V_0$ have to be justified so that the eigenvalues of the matrix become squares of integers, and the sum of those integers equals the total number of valence electrons.

So, given the physical outlook of the molecule in equilibrium, one simply carries out principal component analysis for the matrix $V - V_0$, finding the set of "discrete orbitals", or orbital vectors ψ_i and the corresponding eigenvalues E_i and electron counts λ_i. The elements of the vectors ψ_i reveal around which nuclei the orbital mostly resides; the overlap probability p_{ij} is spatial rather than temporal. For illustration, study the benzene molecule: Benzene is the prototype of aromatic compounds, consisting of six carbon atoms and six hydrogen atoms in a carbon-ring. Altogether there are 30 valence electrons (6 times 4 for carbon, and 6 times 1 for hydrogen). The results are shown in Fig. 8.1 — compare this to Fig. 8.2: It seems that the three first orbitals have essentially the same outlook in both cases. Now there are altogether 7 electrons on the lowest energy level! All orbitals extend over the whole molecule; the hydrogen orbitals are also delocalized — such delocalization applies to all molecules, not only benzene. Note that the orbitals having the same energy levels are not unique, but any orthogonal linear combinations of them can be selected; such behavior is typical to symmetric molecules. The "bonding energy" is the drop in total energy, or the difference between the energies in the molecule as compared to the free atoms; possible values of this energy are discretized, now it (without scaling) is $1 \cdot 7^2 + 2 \cdot 4^2 + 3 \cdot 3^2 + 6 \cdot 1^2 - (6 \cdot 4^2 + 6 \cdot 1^2) = 12$.

The presented approach is general and robust: For example, the unsaturated double and triple bonds as well as aromatic structures are automatically taken care of as the emerging orbitals only depend on the balance distances between nuclei: If the nuclei remain nearer to each other than what is normally the case, there also must exist more electrons around them. Spin considerations are not needed now, as there is no need for external structures (orbitals of "two-only capacity") to keep the system stable and organized. However, no exhaustive testing has been carried out for evaluating the fit with reality. When different molecules were experimented with, the results were not fully satisfactory. Anyhow, the objective here is to illustrate the new horizons there can be available when employing non-centralized model structures.

8.2.3 Characterizing molecules

The time-independent Schrödinger equation (8.3) is not the whole story. As explained, for example, in [10], the complete wave equation consists of *two* parts, the other being time-dependent (and location-independent), these two parts being connected through the energy eigenvalue E. The complete solution has the form

$$\psi(x,t) = \psi(x)\mathrm{e}^{\sqrt{-1}\,2\pi Et/h}. \tag{8.5}$$

Because of the imaginary exponent, the time-independent part oscillates at a frequency that is determined by the energy level of the orbital. Now in the case of discretized orbitals, one can write for the orbital vectors

$$\psi_i(t) = \psi_i \sin \frac{2\pi E_i t}{h}. \tag{8.6}$$

Each energy level also oscillates with unique frequency. This means that the orbitals cannot interact: Because the potentials are assumed to be related to integrals (averages) over the charge fields, there is zero interaction if the fields consist of sinusoids of different frequencies. On the other hand, if the frequencies are equal, the time-dependent part does not affect the results. This way, it seems that each energy level defines an independent interaction mode, and these modes together characterize the molecule — and also each of the individual atoms within the molecule. Thus, define the matrix Ψ where each of the columns represents one of the atoms, from 1 to n, the column elements denoting the

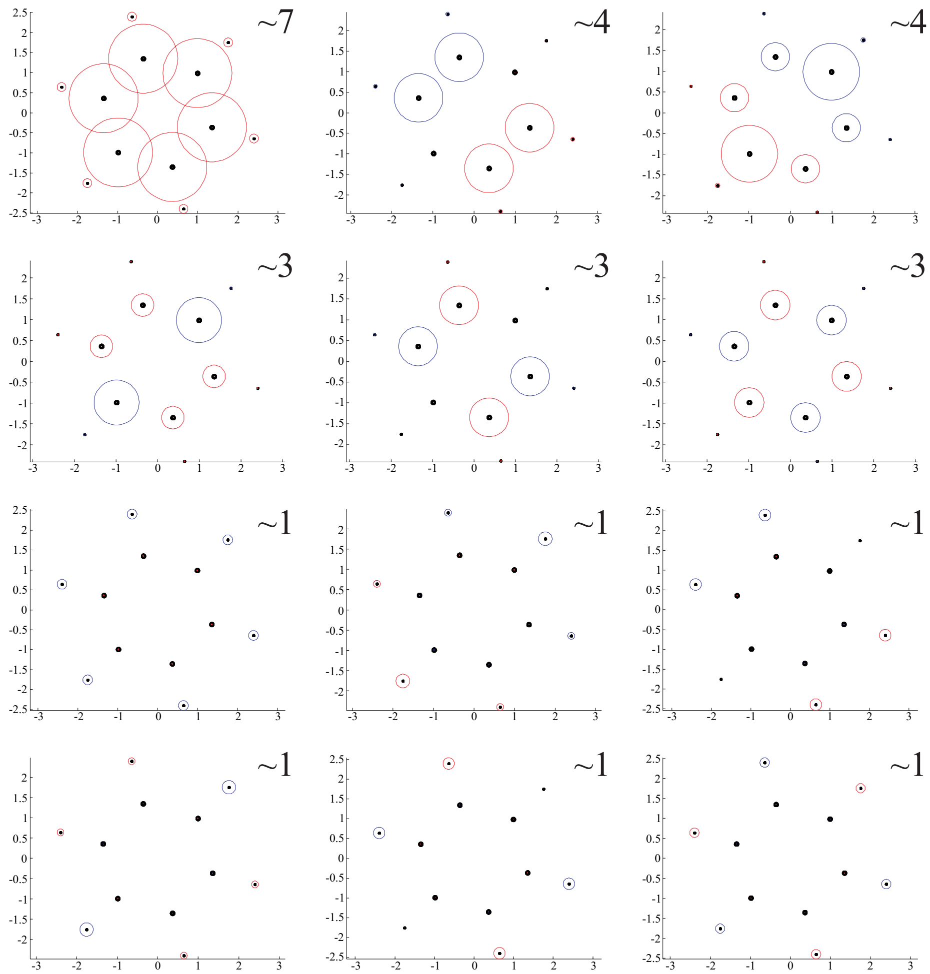


Figure 8.1: "Cybernetic orbitals" ψ_i in the benzene molecule (see text). The larger dots denote carbon nuclei and the smaller ones hydrogen nuclei, distances shown in Ångströms (1 Å$= 10^{-10}$ m). The orbitals, shown as circles about the nuclei, have been scaled by the corresponding λ_i to visualize their relevances. The circle colors (red or blue) illustrate the correlation structures of electron occurrences among the nuclei (the color differences are to be compared only within a single orbital at a time)

contribution of each of the orbitals, from 1 to n, to the total field in that atom:

$$\Psi(t) = \begin{pmatrix} \psi_1^T(t) \\ \vdots \\ \psi_n^T(t) \end{pmatrix} = \left(\; \Psi_1(t) \; \middle| \; \cdots \; \middle| \; \Psi_n(t) \; \right). \tag{8.7}$$

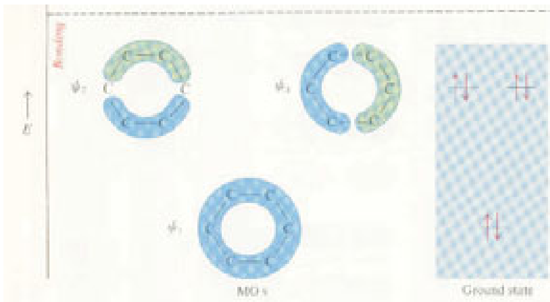


Figure 8.2: Benzene orbitals as proposed in literature (see [56]). Compare to Fig. 8.1

So, rather than characterizing an orbital, Ψ_j represents the properties of an atom j within the molecule. The key point here is that the elements in these vectors reveal the mutual forces between the atoms: If the other of the atoms always has excess field when the other has deficit — the orbitals being "red" and "blue", respectively, as in Fig. 8.1 — the atoms have opposite average occupation by electrons, and the positive attracts the negative. On the other hand, in the inverse case there is repulsion among similar charges. These forces determine whether the atoms can get enough near each other to react; indeed, this force is closely related to the concept of *activation energy* that is needed to overcome the repulsion among atoms. In the adopted framework, this activation energy can be formulated as

$$\Psi_i \Lambda^2 \Psi_j, \tag{8.8}$$

where the total energy is received by weighting the attractive and repulsive components by the appropriate orbital energies (Λ being a diagonal matrix containing the electron counts on the orbitals).

There are only some 100 different atom types, but it seems that there are no bounds for molecule types and behaviors. The above discussion gives guidelines to understanding how this can be explained and how this understanding can be functionalized. A sequential molecule is like a "string" whose vibrations are modulated by the additional "masses" that are attached along it. Because of the linear structure of the protein chain, it is clear that the interaction covariance matrix is diagonally dominant. it is an interesting question how different amino acids in different locations are reflected in the frequency structure of the final molecule. In the proposed framework, it is also possible that the oscillating charge fields interact with the nuclei; this gives rise to extra complexity in the characterization of molecules as the mechanical vibrations can also affect the chemical properties of the compounds — for example, different isotopes having different masses can have differing behaviors.

Because of the universal quantization of the energy levels, the repulsions and attractions are, in principle, comparable among different molecules — but there is the question of synchronization of the oscillating fields. What is more, the proposed characterization of the molecules is not exhaustive: For example, "handedness" is beyond this kind of analysis, and optical isomers have identical representation even though they have differing chemical properties. Because of these shortcomings, the representation is not a general-purpose one — but within a single molecule there may be new tools available.

8.2.4 Folding of proteins and splicing of RNA

All genetic programs are manifested as proteins being products of a complex process of DNA transcription and RNA translation. The proteins are used either as building blocks themselves or as enzymes catalyzing further reactions. Assumedly the proteins are as diverse as the genes themselves — how can proteins have so different properties, being composed of the very basic atoms? It has been proposed that it is the physical outlook, or *folding* of the proteins that is largely responsible for the properties.

The DNA, and after that RNA, only determines the linear sequence of amino acids, the formation of the three-dimensional structures taking place afterwards. But, of course, it is the sequence of amino acids, as being interpreted by the environment, that determines also the final outlook of the molecule: There is affinity among far-apart atoms in the molecule as determined by the structure. Because of its importance, this folding process has been studied extensively, mostly applying computational approaches. But no matter how heavy supercomputation is applied the long-range interactions cannot be revealed or exploited when these long-range effects are abstracted away to begin with in the standard molecular models.

This protein folding seems to be an example of a wider class of phenomena: Intra-molecular affinities have to be understood to master many different kinds of processes. For example, study *RNA splicing.*

In eukaryotic cells, the gene sequences in DNA contain non-coding fractions, or *introns,* in addition to the coding ones, or *exons.* During the processing of pre-mRNA into the actual messenger-RNA, the non-coding portions are excluded in the process of splicing where the exons are connected to form a seamless sequence. What makes this process specially interesting is that the splicing process does not always produce identical messenger-RNA's, but there are alternative ways — sequences can be interpreted as introns or as exons in different environments. It has been observed, for example, that a single mouse gene can in theory produce more different kinds of proteins than what is the size of the whole genome. No doubt understanding the splicing mechanisms will become very important, as nature has found this mechanism because it offers a flexible way to alter the gene expression results without having to go through the highly inefficient route of evolving the whole genome. However, today these mechanisms are still less understood than what protein folding is — and it seems that the real essence of RNA splicing cannot even be explicated yet. Because there is no central control, it is evident that the locations that are to be reconnected need to attract each other. Again, it would be invaluable to master the attractions and repulsions among the atoms in the molecule.

When analyzing reactions, it is often the energy levels before and after a reaction that are studied. However, when studying reaction probabilities, analysis of the final energy levels is not enough: The key point is to see whether the reaction can ever take place. It needs to be recognized that carbon is very reactive, and it forms a bond whenever two atoms are enough near each other — the total energy seems to go down as there are more atoms in the molecule. The most important thing is the activation energy, or the energy that is needed to bring the atoms near each other. Indeed, if the activation energy is low — or, specially, if the reacting components attract each other — the reactions probably take place.

Understanding the underlying principles of attraction and repulsion among atoms gives tools to understand not only the folding processes, but also catalytic (or enzymatic) reactions; and it is enzymes that are responsible of most of the biological reactions. How is it possible that there seems to exist an infinite number of catalysts even though the number of alternative form for "keys" and "locks" seems to be so limited? The new view explains that there can exist an infinite number of energy levels, and thus there can exist an infinite number of attraction patterns. When applying the "holistic" view of molecules as electron systems, orbitals extend over the whole molecule. All atoms count, and it becomes understandable how atom groups far apart can alter the chemical properties of the whole molecule. Specially, in this framework it can be explained how the coordinated-looking very long reaction chains of transcription and translation can exist. The sequential reading of codes can be locally controlled when there is "chained catalysis": Only when the previous piece of code is processed, the next step is catalyzed by the previous reaction result.

Now there are enough tools to implement an "emergent-level simulation" for modeling the protein folding: Rather than doing the extensive quantum theoretical ab-initio simulations, represent the chemicals in terms of their emergent affinity structures Ψ, and shuffle the potion, producing a distribution of chemicals and their structures in equilibrium. That is, put in the "codes" — linear amino acid sequences — and let the environment interpret and process them into a "system". Note that the representations Ψ change as the physical appearance changes, the atoms traversing in the force fields, and these also need to be adapted; what is more the physical outlook also affects the possibilities of the atoms to approach each other, and this also needs to be taken care of in the simulations. Whenever orbitals come near enough each other, they automatically merge, thus forming a bond[1]. The simulation is also simulation of nuclei movements: Given a nucleus combination, the orbitals can be determined, and after the charges are found around the nuclei, and when the physical constraints are taken into account, the total forces affecting the nuclei can be calculated. When the nuclei are moved slightly in the directions of their attraction, the orbitals need to be calculated again, and the whole loop is repeated, until a balance is reached so that no residue attraction remains.

The interpretation of one-dimensional code into a high-dimensional operational representation (see 7.3) is very literal in the case of protein folding — the chain of amino acids becomes a many-folded structure. The end result is the folded protein; and it has been claimed that it is this structure that mainly determines the functions of the molecule. Is there anything more to be said here?

After all, one is interested in the functions, the tissues or catalytic effects of enzymes, not in structures themselves. Is there any possibility of analyzing how the function of a folded protein emerges from its structure in general terms? It should be clear that it cannot be the physical structure alone that would offer the complete answer: Molecules cannot perceive physical dimensions; neither is it some three-dimensional jigsaw puzzle in the tissues. It is tempting to

[1] As soon as there is *any* interaction among atoms, the potential matrix has non-zero non-diagonal elements, and a mixture of orbitals is found; indeed, this model predicts the *tunneling* of electrons also among far-apart atoms

hypothesize that it is not only the physical structure, but again the cybernetic structure that plays a role here: As the energy levels of the molecule specify its oscillatory structure in the quantum level, perhaps neighboring molecules find synchronization and become aligned. There emerges resonance, and the molecule-level structure is repeated and magnified, being manifested as a special type of homogeneous tissue, or — why not — as a crystal lattice in the inorganic case. The resonances define a Pythagorean "Harmony of the Phenospheres", cybernetic balance of vibrations ... is this the key to the next emergent level where the molecular components are combined into a next-level structures?

8.3 Towards "cosmic cybernetics"?

The previous example raises the question whether the physical systems also are cybernetic. If the cybernetic principles are rooted so deep in the structure of matter, perhaps there are reflections also on the wider scale? After all, the systems in the large also are locally controlled only, and there is balance and non-trivial structures in the universe. The advocates of "theories of everything" claim that after combining the different kinds of basic forces, everything can be explained; however, problems of cognition, for example, do not belong to this "everything". The real GUT theory has to explain emergence; does the cybernetic thinking have something to offer here?

8.3.1 Formation of stellar structures

For example, when looking at complex physical systems, like galaxies, etc., where there also are individual "agents", the stars, one can see that together the bodies constitute seemingly self-organized, long-living constellations. As compared to the neocybernetic studies, and when studying the possibilities of extending such considerations to stellar systems, there are many problems. First, because of the geometries and the outlook of the gravity law, the formulas are highly nonlinear; second, as the forces are determined by the mutual orientations among the stars, and as these orientations constantly change, it seems that there cannot cumulate any structure among the stellar bodies. The main problem, however, is that there seems not to exist a repulsive force: Gravity attracts bodies, and if there exist no balancing mechanisms, no cybernetic self-regulation or self-organization can emerge.

Still, the solar systems and galaxies *are* rather stable. Originally there were only clouds of stellar dust, but, based on local interactions only, different kinds of structures have emerged. There evidently exist balancing effects — and, indeed, as the attraction increases speed, the experienced centrifugal forces try to drive the bodies apart.

How can the cybernetic ideas of elasticity, optimality, and adaptation be motivated when it is mindless stellar bodies that implement all functionalities? However counterintuitive it may sound, all mechanical systems in the global scale for some reason implement optimization. In *Lagrangian mechanics,* and later in *Hamiltonian mechanics,* it is observed that the Newtonian laws of motion can be reformulated as optimization problems: Along the motion trajectory, the time integral of the quantity $L = W_{\mathrm{kin}} W_{\mathrm{pot}}$ reaches its minimum value, where W_{kin} is the kinetic energy of the system, and W_{pot} is the potential energy. This starting point has been applied a long time for deriving dynamic models for mechanical systems, but it is not applicable for analysis from the point of view of neocybernetics.

Since the Newtonian times, the emphasis in dynamic modeling has been on accelerations induced by forces. However, in a galaxy in balance, such accelerations — even though everything is after all based on them — seem to be rather irrelevant. As in the case of electrons before, now try to eliminate the accelerations. Acceleration means change in velocity; assume that the velocities remain constant in the large, and the neocybernetic steady state prevails. Study the possibilities of finding an "emergent model" for a galaxy where the component-level interactions are abstracted away. What would such a model perhaps look like? What are the local-level adaptations that can be justified in such a non-adaptive environment where the laws of mechanics and gravity cannot be adjusted?

What kind of alternative "stationary" models there are for central motion? A characteristic model family is offered by the Navier-Stokes equations of fluid dynamics: Vortex structures in flows are commonplace. Indeed, the vortex model offers intuitively appropriate behavioral patterns also for representing gravitational fields, down until the black hole like singularities. Correspondingly, in an combined electric/magnetic field such rotors also exist. The common denominator characterizing such models are *vector products* between some vector-form quantities. Such mathematical representations can be seen as the *emergent evolutionary goals* that are easier visible in systems with vortices of faster time scales than in slow gravitational systems that still are in their transitory state towards that final balance. In each case, the underlying actors (photons vs. mass units) assumedly only obey tensions determined by their local environments, and the challenge now is to reinterpret the variables to implement the observed behaviors in terms of physically relevant, locally observable quantities.

Of course, this all is very vague, but some bold (yet schematic) conclusions are here possible. The hypothesis here is that the *eventual stationary gravitational vortices can be modeled using vector products between the velocities of bodies and the forces acting on them.* One motivation for this selection is that the product of (scalar) velocity and (scalar) force has the dimension of power — and this quantity sounds like a reasonable emergy measure to be pursued in a mechanical domain. Assuming that the momentary velocity vector of a particle j is denoted as v_j and the total force acting upon it is F_j, the size of the corresponding vector product is $|v_j|\,|F_j|\sin(\alpha_j)$, where α_j is the angle between the vectors, and the average of it is $\mathrm{E}\{|v_j|\,|F_j|\sin(\alpha_j)\}$. When approaching the steady state, the average gets nearer to the momentary value.

But what is the assumed adaptation scheme among mindless bodies, why should they try to do maximization of some quantity and how could they implement that? It needs to be observed that orbits that maximize the cross product criterion (under converged central motion) are *circular orbits* where the velocity and acceleration remain perpendicular. It can be assumed that if such circularity constraint is fulfilled by all stellar bodies, their *orbits minimally interact* and remain intact, whereas non-circular behaviors are more probable to die out through collisions. This means that circularity is evolutionarily beneficial: Circularity of orbits serves as an emergent fitness criterion — even though the actors never see the chaos around them in such a wider perspective.

If the gravitational system is interpreted as an elastic one (stronger force meaning shorter radius and faster motion when the angular momentum is preserved), the product $|F_j|\sin(\alpha_j)$ can be seen as the external input, and $|v_j|$ can be seen as an internal variable, one can express the local-level aspirations in the neocybernetic framework when one defines the system state as $x_i = |v_j|$ and the inputs as $u_j = |F_j|\sin(\alpha_j)$. When the local maximization takes place, global-level emergence assumedly takes place: The final state after convergence under elasticity assumption minimizes the familiar cost criterion

$$\min_x \left\{ \mathrm{E}\{\frac{1}{2}x^T \mathrm{E}\left\{xx^T\right\} x - x^T \mathrm{E}\left\{xu^T\right\} u\} \right\}. \tag{8.9}$$

Looking closer at (8.9), one can see that the former term represents *kinetic energy,* being essentially based on products of velocities; the diagonal elements represent translational components, whereas the non-diagonal entries are inertial components. Following the neocybernetic spirit, the state vector x can be compressed in the PCA spirit so that the essence of the system is minimally affected. It is here where the added value can be seen: As the state space is compressed, the huge number of individual mass entities becomes represented by a simpler model. This essence of the emergent model is in the lower-dimensional *inertia matrix* that can be determined through observations as $\mathrm{E}\left\{\mathrm{xx^T}\right\}$. It turns out that *the uncoordinated stellar units can be seen as a more or less three-dimensional rigid body.* Still, before the final convergence in infinity, there is still noise in the system: The local whirls among the stellar bodies are manifested in the model as non-vanishing variations in the data. In principle, the model does not only capture the movements of the galaxies and stars, but also stars and planets, and planets and moons in its statistical structure. The larger model can also be decomposed: Single solar systems can be characterized in terms of their local PCA structures. The evolution towards circular orbits around a single mass center still continues — for example, the tidal effects gradually bind the motions of the moons.

Perhaps the Pythagorean "harmony of the celestial spheres" can be defined in terms of principal component modes (stars on the wider scale and planets on the narrower one), modulating dissonances being caused by local rotations (orbiting planets and moons, respectively, etc.) and ellipticity of the orbits.

8.3.2 Everything, and more

The neocybernetic principles can be claimed to span behaviors from the elementary level (orbitals) to the cosmic level (orbits) — but hypotheses can be made even beyond that, towards the most extensive levels of all.

The neocybernetic models pop up in very different cases: Perhaps this is not a coincidence. Many complex systems can be characterized in terms of optimization, models being derived *applying calculus of variations,* the explicit formulas (constraints) being the emergent outcomes of underlying tensions. For example, *Maxwell's electrodynamics* can be formulated in terms of such optimization.

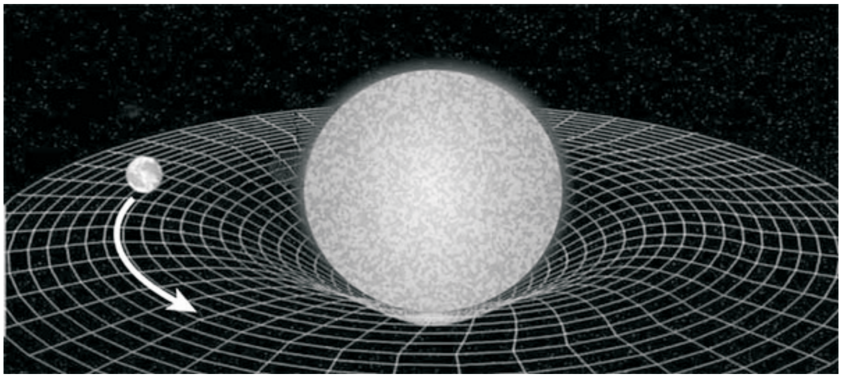


Figure 8.3: Illustration of how the four-dimensional space-time is deformed as a concentration of energy (mass) is found in the universe

What if the idea of distributedness of underlying stochastically behaving actors is explicitly employed, perhaps it is the neocybernetic framework that automatically pops up?

The faith here is that all observable behaviors even in the most universal levels are finally implemented by uncoordinated low-level actions. In addition to the stellar and planetary motions, there are also connections between astrophysics and cybernetics on the more fundamental, less immediate level. It seems that all non-trivial systems that still exist after those millions of years have found their ways of implementing adaptive feedback, otherwise no self-regulation and self-organization could have been reached. And, indeed, it seems that the idea of elasticity is intuitively applicable also in wider scales — for example, the Einsteinian tensors are tools to formalize pressures, or tensions (see Fig. 8.3). There are problems when trying to apply the relativity theory to elementary particle level; in this scale, all variables are quantized and the continuum models collapse. There also exist models based on “cell-structured universes” that are (more or less) compatible with the Einsteinian cosmology — the cybernetic ideas could directly be applied in such models, neighboring cells being interacting subsystems transferring emergy: Compare to Fig. 3.5, where impulses traverse through the “space” as the coupling between “cells” is tight enough.

There are efforts to find the Grand Unifying Theory (GUT) that would combine all basic forces like combine electromagnetics, gravity, weak and strong nuclear forces into the sme framework. Elasticity seems to offer fresh ideas also in the field of basic physics: Beyond the observations, in super string theories, the elementary particles are seen as vibrating springs (or *vibrations* of strings). But regardless of the form of the final theories, it seems that thinking of the universe as an elastic self-balanced shell reacting to external pressures, this “shell” being distributed in matter particles, offers a useful framework for studying matter. The Heisenbergian thinking is to be extended, as it is all interactions (not only measurements) that affect the system, the effective variables being reflections of the emergent balance among the system and the environment. Measurable variables are “interaction channels”, each interaction mechanism introducing a spring of its own; individual seemingly static relations of the form $\bar{x}_i = q_i \bar{u}_i$ connect observations through some coupling coefficient q_i. The natural constants are not predetermined, but they are the visible manifestation of the balance ratio between reaction and action. The modern theories employ some 11 dimensions (some theories necessitating introduction of several dozen dimensions!) where there are some "collapsed dimensions" among them: Now there is no upper limit to the dimensions as they are no actual coordinate axes but they only represent the number of interaction channels into the universe; and it is easy to think of the vanishing degrees of freedom as being only tightly coupled to others through the cybernetic feedback controls. The constants of physics should not be seen as predetermined quantities: There are propositions that the natural constants are gradually changing as the universe gets older. One of such propositions is by Paul Dirac, who claims that cosmology should be based on some dimensionless ratios of constants (known as "large number hypothesis).

If the cybernetic thinking universally applies, one can exploit the understanding concerning such systems: Perhaps universe as a whole is *optimizing* some criterion? This would help to escape some deadlocks one is facing today.

It has been estimated that to have a stable, non-trivial and long-living universe that can maintain life, the natural constants have to be tuned with $1/10^{55}$ accuracy. Such astonishing coincidence has to be explained somehow, and different families of theories have been proposed. First, there are the *anthropic* theories, where it is observed that the world just has to be as it is, otherwise we would not be there to observe it; the other theories are based on the idea of *multiverses,* where it is assumed that there is an infinite number of "proto-universes" in addition to our own where physics is different. However, in each case it seems that physics reduces to metaphysics, where there are never verifiable or falsifiable hypotheses.

If the universe is (neo)cybernetic, each particle maximizes the share of power it receives, resulting in the whole universe becoming structured according to the incoming energy flows. Then there is no need for multiverses, as it is the only the best alternative that really incarnates. It is as it is with simple subsystems: Fermat's principle says that light beams "optimize" selecting the fastest route; it is the group speed that determines the wave propagation, the emerging behavior representing the statistically most relevant alternative. Similarly, the only realized universe is where the optimality of energy transfer is reached.

Again, applying the neocybernetic intuitions concerning adaptive controls (as studied in chapter 5), the reasonably evolving universe must not be too balanced: Perhaps there must exist some level of asymmetry to avoid the final stagnation and subsequent collapse — this collapse perhaps signaling the end of the universe as we know it, giving room for the "next version" to start the cycle from the beginning.

The idea of the evolutionary universe is intriguing. How about the adaptation and evolution mechanisms? The key point is not whether the cybernetic thinking can be applied to modeling non-living physical systems — the most interesting question is whether the universe can be interpreted as being a *living entity* itself. Perhaps this all is not only loose metaphysics: As studied in the next chapter, it seems that there seems to exist a nice connection between the universal physical principles and the neocybernetic systems.

Level 9

Arrow of Entropy and *Origin of Life*

As was observed before, in some cases there are credibility problems when trying to model complex systems applying simple methods. However the credibility problems are fixed, at some stage the models become not only incredible but truly impossible. When very improbable phenomena are assumed to cumulate *ad infinitum,* no time in the universe is enough to make emergence to happen.

There exist plenty of examples of such very improbable processes in biology: For example, the gene transcription from DNA to RNA consists of a huge number of marvelous coordinated-looking steps that are needed, and so does the translation process of RNA further to proteins. How do the locally controlled atoms know when to adhere and when to let loose when the sequential reading of the codes is being carried out? And these are only subprocesses — above them, there are the developmental processes in an individual and evolutionary processes in a population that are equally astonishing. How do the systems climb the endless steps of increasing complexity?

Admitting that there are still challenges is the first step towards more plausible models. Unprejudiced analyses make it possible to see things in a perspective — and, suddenly, it turns out that all is clear. When correct interpretations are applied, it turns out that actually the systems are not going in the direction of increasing improbability; they go down towards *maximum* probability. The systems struggling against the flow of entropy is just an illusion (see Fig. 9.1).

In this chapter, this viewpoint is applied to the analysis of how life could have emerged from the non-living. Indeed, it does not matter how long the ladders are; when you are going in the right direction it does not matter how long it takes. There is enough time — as long as the processes go in the right direction.

9.1 Thermodynamic view of cybernetics

The most universal framework that governs all physical systems is *thermodynamics.* The thermodynamic concept of *entropy* is among the most fundamental

Figure 9.1: The paradox of entropy flow is just an illusion (graphics by Maurits Escher)

ones in nature, and when searching for universal laws governing cybernetic systems, among others, these issues need to be addressed.

9.1.1 Entropy and order

Applying the thermodynamic interpretation (as defined by Rudolf Clausius), entropy reveals the extent to which the energy in a closed system is available to do work (as defined in a somewhat sloppy manner). The lower the entropy level is, the more there is *free energy.* In a closed system, entropy level cannot decrease; it remains constant only if all processes within the system are reversible. However, because the natural processes typically are irreversible, entropy in the system increases, so that energy becomes "inert". Even though the total amount of energy remains constant, according to the *first law of thermodynamics,* it becomes less useful, according to the *second law of thermodynamics.* Ultimately, the system ends in a thermodynamic balance, or "heat death", where there is no more free energy available.

This direction of increasing entropy seems to be opposite to what takes place in cybernetic systems. The accumulation of complexity in the evolving structures seems to fight against the second law of thermodynamics. The easy answer here, of course, is that cybernetic systems are *open systems,* where there is energy transfer between the system and its environment — the total entropy level in the whole universe increases despite some "countercurrents" in the flow. However, there are some rough edges in this explanation: This assumption means that the strongest of theories, thermodynamics that should govern everything, is not applicable in cybernetic systems, becoming void and useless *if the system just decides to develop.* There seems to exist a gap between "normal" and "abnormal", evolving systems (see Fig. 9.2). Why do not the cybernetic systems not choose the "easy way", following the flow of entropy?

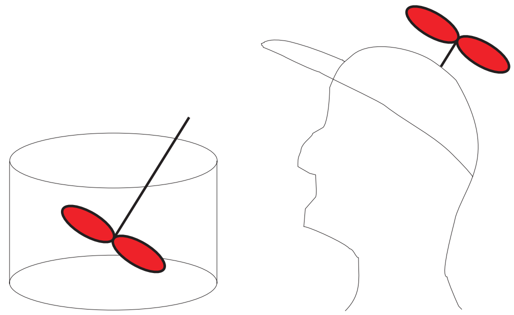


Figure 9.2: Two vessels — an ideal mixer and an “idea mixer”. Two systems where there seemingly is something very different from the thermodynamic point of view: In the former case, in the perfectly stirred tank, addition of energy decreases order and structure, whereas in the latter case, in the cognitive system, activity increases order, new structures being constructed

There also exist different, more or less closely related definitions to entropy. In statistical mechanics (by Ludwig Boltzmann and Willard Gibbs), and analogously in information theory (by Claude Shannon), entropy is related to probability: More probable states (observations) reflect higher entropy than less probable ones. In a sense, entropy is the opposite of information — less probable observations contain more information about the system state. In such discussions, the second law of thermodynamics, or the increase in entropy, is reflected so that systems tend to become less ordered, and information becomes wasted.

It seems that intuitions concerning entropy are to some extent contradictory, or at least obscure. One hypothesis assumes that entropy, being among the only one-directional quantities in physics — defines the direction of time. Perhaps the most marvelous conclusion is that the universe cannot shrink because that would make particles be closer to each other — thus the system being more ordered, total entropy in the universe going down. This would also mean that time would start going backwards! Perhaps there is room for yet other interpretations.

The probability-bound interpretation of entropy is appealing, but it also seems to result in paradoxes: For example, a symmetrical pattern is intuitively more ordered, containing more information, and consequently having lower entropy than a completely random pattern — on the other hand, symmetric pattern can be seen to contain *less information* than a random pattern, because the redundancies caused by the symmetricity can be utilized to represent the patterns more efficiently, so that the entropy level should be *higher* now. Indeed, it can be claimed that the “algorithmic entropy” is higher in a symmetric pattern than in a non-symmetric one. To confuse concepts concerning order and symmetry even more, or, rather, to reveal the inconsistencies in our intuitions, think of

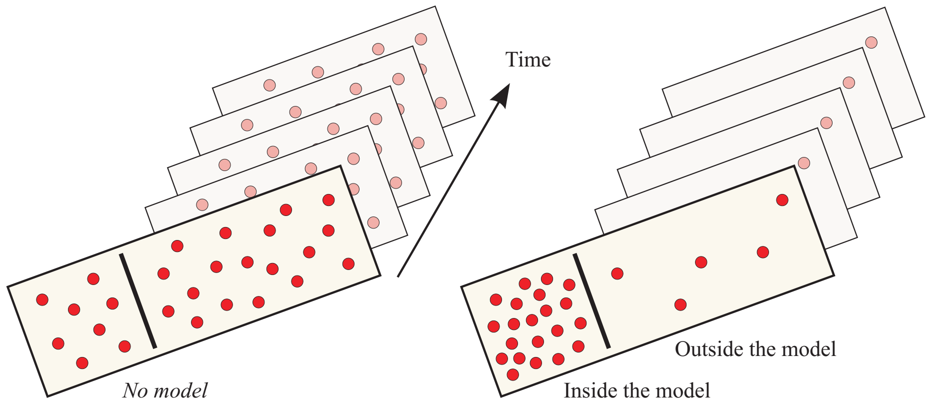


Figure 9.3: Illustrating the effects of cybernetic control (dots denote "information units" and their conglomeration): The case *without* model (on the left) has high probability at start time, and thus high *momentary* entropy; the case *with* model (on the right) has low total information over time, and thus high *sustainable flow* of entropy. Note that the information stored in the model is constant and thus negligible, being defined once for all

the following claim: A totally unordered system can be said to be extremely symmetric as the components cannot be distinguished from each other.

The intuition has been seen as the basic tool in neocybernetics to reach good models — such thinking seems to collapse here; but can the power of intuitions still be preserved? The answer is yes — here it is assumed, according to the original intuition, that orderliness, or loss of disorder, is a manifestation of low entropy. The key point here is that the simplicity of symmetric patterns, or ordered patterns in general (loss of information in them), is just an illusion: The missing information of the pattern is buried in our mental pattern recognition capability. If the same data is to be presented without the supporting underlying cognitive machinery, or specialized interpretation and analysis tools, there is no handicap — the redundancy cannot be exploited, and no compression of data can be reached. In general, a higher-level representation makes it possible to abstract the domain area data; in other words, as has been observed, a model is the key to a compressed representation. And this idea can be extended to cybernetic systems in general: It need not be our personal cognition machinery that constructs the model storing the excess information; any cybernetic system can do that in its own more or less narrow environment.

In the cybernetic perspective, the two views of entropy can be combined in a natural way: On the one hand, it is about balances and pursuit towards heat death, and, on the other, it is pursuit towards least information, as measured in terms of variation. What is more, it turns out that *evolution of structures increases entropy.*

9.1.2 Control changes it all

The cybernetic systems, as studied before, are characterized by balances: First, the determination of $\bar{x}$ is based on finding the dynamic equilibrium as determined by the system model. Second, the system structure as determined by the matrix $\mathrm{E}\{\bar{x}\bar{u}^T\}$, is also a dynamic equilibrium as determined by the statistical properties of the environment. Indeed, in a cybernetic system there are balances at each level — and, in this sense, the convergence towards a steady-state model is completely in line with the second law of thermodynamics: Variation (information) around the balance is maximally being eliminated, the "local heat death" almost being reached.

Where does this balancing property come from? It is the structure and order on the higher level — or the model — that makes it possible to control the lower level, or to reach the information elimination there. Evolution is the process of introducing ever more complicated structures that facilitate ever better control of the environment, either implicitly, as in lower-level biological systems, or explicitly, as in man-made systems. In any case, the cybernetic controls boost entropy — and the more sophisticated the control is, the higher is the rate of entropy production. In this way, rather than opposing entropy, the cybernetic system tries to maximize entropy — quite in accordance with normal physical systems. It is all about correct viewpoint, and selecting the system boundaries appropriately. It is the control system intuition that is needed to solve the "arrow of entropy" paradox.

Because of the simple definition of information (information being manifested as variation), it is possible to distinguish between information being captured in the structure (the model) and information being left in the signals (unmodeled noise). The cybernetic system acts like a Maxwell Daemon, distinguishing between two "containers" of information and noise, compressing information and pumping "negative entropy" into the emerging structures, thus causing positive entropy be left outside the structures (see Fig. 9.3). The key point here is that the single container of negative entropy (the model) is outweighed by the large number of samples with increased entropy level (data variation in the environment being suppressed), thus being a thermodynamically sustainable scheme. Whereas the momentary entropy increases, the "emergent entropy", the average of entropies over the whole environment and over all future decreases when the cybernetic strategy is employed. The same thinking applies to entropy as to other quantities when the neocybernetic perspective is applied: The time axis is abstracted away, only the average over the long-term loss of information is considered.

Perhaps the most important consequence of the new interpretation of the cybernetic systems is that reductionistic approaches become possible: Traditionally, the only systemically consistent level of studying such complex systems, with the whole environment being involved, was the holistic level, the whole universe being seen as one entity. Now each subsystem can be studied independently, as an independent thermodynamically consistent entity. The traditional view of seeing the relationship between the system and the environment is turned upside-down, or actually inside-out, environment now being the innermost part, being controlled by the system (see Fig. 9.4). It is the environment that is seen as the object, and the system is the subject, manipulating the environment, and a

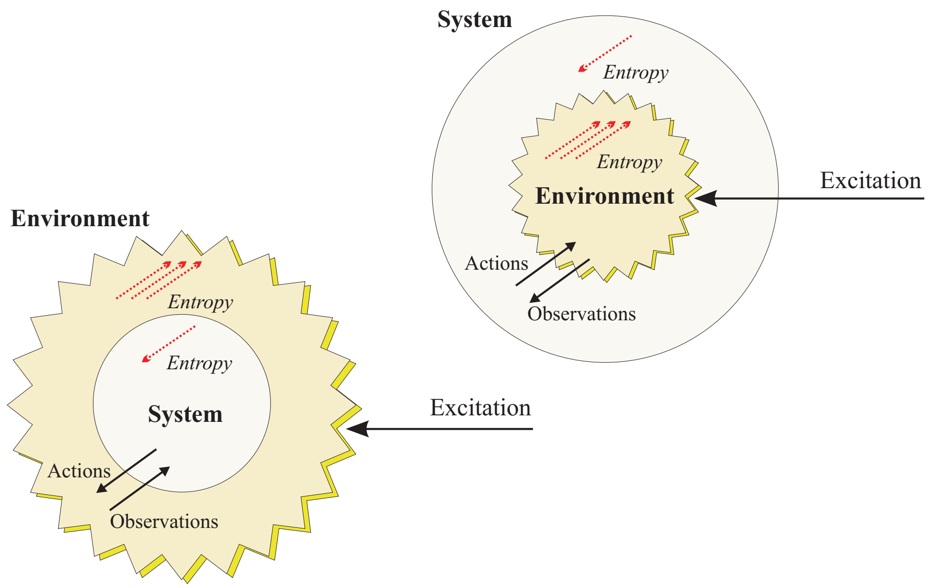


Figure 9.4: Reaching entropic consistency within systems. Left — traditional view, right — cybernetic view

more complicated system always sees the lower levels as through a looking-class. The highest-level model where the negative entropy is concentrated remains outside the boundaries. The original input into the environment is white noise; as seen by the highest-level system, the lower-level systems distort this noise, and the systems tries to capture this distortion, or the redundancy there is in the observations. It does not matter how many levels of systems there are, the same principle of modeling always applies. Note that the low-level systems only see a narrow view of the complex environment, and it is only this limited information that is relevant to that system — the whole complexity of the world needs not be captured.

9.1.3 Another view at model hierarchies

When the view of "information units" is employed, it is perhaps motivated to take a closer look at models themselves. The model is a container of information that characterizes the patterns that distinguish the system in question. To have some perspective, note that the cybernetists Norbert Wiener and Arturo Rosenblatt have argued that

> The best material model of a *cat* is another, or preferably the same, cat.

However, this view only applies to a trivial structure of models, when just a isolated single cat is being analyzed. In the beginning, modeling truly starts from representing all available information, the "world model" consisting of the data directly, but when abstracting over individuals, the model becomes more compact, the set of common patterns becoming smaller. When consistent vari-

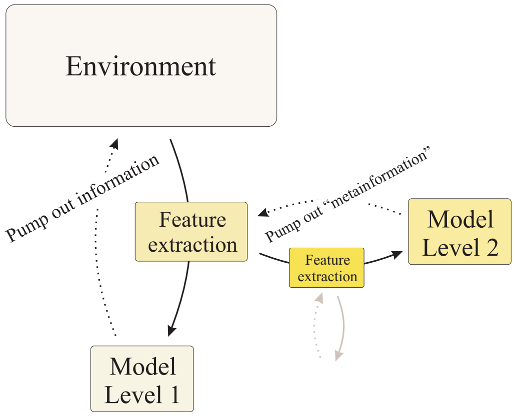


Figure 9.5: The challenges in "universal neocybernetics" are compressed into the question: How are the domain-specific features (pieces of information) extracted from environment

ations from the nominal are detected, the best model for the whole category of cats summarizes the similarities or invariances characterizing all of the "cat" samples. Neocybernetic models are the collections of *invariances over variances.* This model compression, or separation of information between the containers is implemented automatically by the cybernetic adaptation and control mechanisms. At this level, the model contains the detected *similarities.*

But this kind of simplification is not all that happens. As has been observed, neocybernetic systems are not alone — and, similarly, also neocybernetic models form a hierarchic interlinked structure. If there is a hierarchy of models, further compression of the sub-level model takes place: Redundancy in information gives rise to a higher-level "model of models", or, as seen in the control perspective, "control of controllers". A generalized view of control can be based on the view of *eliminating information invariants,* or transferring them onto the higher level, being shared by different domains. As there are common patterns among features, the higher-level model captures this redundancy. This means that the same information is represented in the lower-level model *only once.* At the higher level, when differentiating submodels from each other, the model contains the detected *differences.* The hierarchy of cybernetic models optimizes among the representations of similarities and differences, assumedly minimizing the size of the overall model.

Seeing information as bits — in the spirit of information theory — makes different levels of controllers commeasurable. No matter how a feature is defined, directly in terms of a measurement or through a complicated algorithm, there is no qualitative leap in their algorithmic complexities; information can be collapsed onto the same format, and structures within the models also become a matter of analysis and control. At the lower level, there is identity among information sources (features) that deliver the same information (distinguishing between categories in an identical way), whereas when seen from above, the information (algorithmic complexity) in the model is minimized, so that the simplest representation remains. The cybernetic adaptation and compression of information always follows the same principles (as studied in chapter 3) — the key issue when escaping in a phenosphere to a higher level is that of determining the features (see Fig. 9.5).

This analysis that is based on the formalized view of information applies also to memetic systems — for example, when looking at *science,* one can even find new perspectives into the groundings of model thinking. Simplification of a model is a manifestation of existence of a higher-level control, and when studying science, these models are *theories,* also becoming more compact as being "controlled" by the higher-level controls; these controls are defined by the paradigm of doing sciences in general, governing the principles of all scientific work. This means that as any science is a subsystem in a controlled hierarchy of cybernetic sciences, it is bound to become more and more simplified as the hierarchy matures: Only the most powerful explanations survive. Indeed, such simplicity pursuit (compare to "Ockham's razor") is traditionally taken as the philosophical foundation of all modeling without any attempt to justify it. As the scientific discipline is a cybernetic system as is the subject of its study, this simplification is perhaps not only an engineering-like shortcut: The same kind of simplification takes place also in nature.

The qualitatively separate levels in the models are also visible in practical real-life systems — study the flows and information hierarchies in an industrial process plant:

1. **Physical flows** are the real flows of matter and energy in the process.
2. **Information flows** typically consist of the feedback controls governing the physical flows.
3. **"Knowhowflows"** consist of supervision and optimization of the underlying control structures.

The goal of traditional control is balancing of the time-domain dynamics by exploiting the causalities; this process-specific layer supplies for the cybernetic medium to be exploited by the domain-independent cybernetic structures. Minimization of variances in product quality, and robustness against environmental disturbances, is implemented finally applying the concrete controllers. The ideas of feedback control are the same in all kinds of physical systems. In the similar way, the ideas of cybernetics are still more general, covering all kinds of control systems, abstract or concrete. One has to proceed from "bulk information" to *metainformation* or knowledge (information on information). The level of "cybernetic controls" on the metainformation level is somewhat ill-defined — indeed, as soon as all information flows in a cybernetic loop are unambiguously fixed, it becomes a traditional control loop. Even though (as observed above) one always operates on the same kind of information units, it is reasonable to distinguish between levels: The cybernetic framework combines systems from different phenospheres. The plant-level idea of an industrial system is functional, combining subsystems, both memetic and physical in appropriate, ingenious ways. Information is transferred between phenospheres, or "parallel universes of information"; one could distinguish between *eksoinformation,* or "inter-domain information", and *endoinformation,* or "intra-domain information". The lower-level controls exploit the available endoinformation that can be included in the formal loops, whereas the eksoinformation can be called "knowhowflow". In knowhowflow one can exploit expert understanding and common sense reasoning, and humans are integrated in such closed loops. This flow is typically very stochastic, and there are only the tensions visible; it is impossible to formulate the actual processes explicitly, but, again, the final state is well-defined. The key point in expertise exploitation is selection and preprocessing of the appropriate variables and weighting of them. The drifts are manifested in the engineering-like pursuit towards better solutions — cheaper, faster and more accurate measurements, actuators, and algorithms. Whenever the features are formulated, or the "probes" are defined (see chapter 4), model adaptation takes place in the familiar way. There is a balance among technological possibilities and economical constraints.

9.1.4 Principle of maximum entropy production

Traditionally, the second law of thermodynamics is thought of as being a universal, more or less metaphorical principle. The existence of systems with inverted, entropy-decaying nature has made it difficult to motivate explicit utilization of this principle in practice: It seems that the entropy principle cannot be applied in a reductionistic way for analysis of concrete large-scale systems.

Now, according to the above discussions, the entropy in a subsystem always increases when seen from the higher-level system. In a cybernetic system, entropy increases in a consistent manner, there is "balance pursuit" at all levels, completely in line with the second law where thermodynamical balance is the ultimate goal. Because of this consistency, any subsystem at any level — when its boundaries are appropriately determined — can be studied separately, and also holistic systems can be analyzed in a reductionistic manner. In this sense there is no more difference between different kinds of complex systems: Living systems and non-living ones, for example, can be modeled in the same framework. Whereas the first law of thermodynamics (energy principle) offers powerful tools for deriving static models, it seems that the second law (entropy principle), being a fundamentally flux-based concept, offers generic tools for deriving *dynamic models* — also for complex adaptation processes. As long as there are in a thermodynamic system heat resources, or heat differences, there is capacity to do work; similrly, as long as there is emergy in a cybernetic system, there is capacity to adapt and "live". There is directed (generalized) diffusion, or "leakage" of emergy from the environment, evolution making this leakage from the reservoirs become faster.

If entropy production can be seen as a consistent process, the next step is to assume that it happens *as fast as possible.* It can be assumed that it is the most efficient strategies that only remain visible, characterizing the whole system when seen from outside. In the spirit of principles of *least action* or *minimum energy,* as originally proposed by Maupertuis, and later extended by Euler, Lagrange, and Hamilton, one can propose the *principle of maximum entropy production* for characterizing the processes of information decay. Such somewhat teleological modeling principle give strong tools for looking systems and their adaptations in a perspective.

When information is seen as a concrete measurable quantity, formally incompatible systems can be put in the same framework, and the intuitive visions concerning behaviors in cybernetic systems can be functionalized. The entropy levels, or, rather, changes in the levels, determine the *free energy,* and they can be applied as a measure for tensions in a cybernetic system; this measure can be expressed as bits of information. Let us study what this means in practice when doing entropy pursuit — what is the maximum speed of information container separation, or what is the rate of the "emergent dynamics" in cybernetic adaptation processes?

Assume there are k samples of data. Information is extracted from this data in terms of more or less computational features, defining different ways of looking at the system. In the spirit of information theory, the features are reduced to one bit: They contain elementary characterizations of the form "yes" or "no". Further, for simplicity assume that the probability that a random independent feature gives the correct classification is p; then the probability that the feature remains indistinguishable from the others all the time during sampling is

$$p^k. \tag{9.1}$$

If the acquired information is optimally exploited, the probability that a superfluous feature does not become ripped off decays exponentially. In practice, the optimum speed of adaptation becomes dependent of the signal-to-noise ratio in data — but at last in this (extremely) simplified case, the rate of elimination of bits in both environment and in the model takes place exponentially. Indeed, this is what one would expect.

The above discussions are not only a theoretical exercise: They offer powerful conceptual tools to attack complex evolutionary systems. The new view turns the direction of thermodynamic tensions, changing the destructive-looking tendencies into constructive ones, making the originally improbable developments probable after all. The power of the new intuition is illustrated here by discussing one of the biggest mysteries there is — the *origin of life.* The question about the origin of life is not only a philosophical problem: When trying to extend biology from the analysis of distinct examples, individual animals or species, to the scientific analysis of general principles, it is necessary to understand the common principles all of them share: Emergence of a cybernetic system is a birth of an individual. Fundamentally, it is this origin of life that is faced by each living system — after all, the individuals repeat this process in their development starting from non-living chemicals.

9.2 Ladders towards life

The processes of DNA transcription and translation into proteins are much too amazing to be credible — and still it happens all the time. It seems so cleverly orchestrated that it is easy to assume that a guiding hand is necessary in these processes. Thinking of this all having emerged all by itself, by some "blind watchmaker" — this is the basic dilemma where creationists have all the memetic weapons at their hands. However, in the cybernetic setting the things turn upside down, and it can be claimed that *the most plausible explanation for origin of life is non-divine.*

9.2.1 Paradoxes of living systems

There exist various seemingly reasonable presentations about the origin of life in the literature — for example, see [13] and [23]. There seem to exist no real problems whatsoever, and even questioning the trivializations is seen as next to insane [21]. However, it seems that it is these explanations that are missing the common sense.

When the origin of life is discussed, it is often claimed that the key problem is to explain where the first DNA molecule came from — after that, reproduction etc. should be no more problem. And when a reproduction machinery is available, and mutations in the code take place, it is the Darwinian principles that only need to be followed to explain today's diversity in nature. Even though such emergence of DNA is highly improbable, there was billions of years time. Unfortunately, as revealed by the famous Miller-Urey experiments, just adding mindless energy in the potion of chemicals, only simple amino acids are produced — more complex molecules just are not energetically stable enough. But it is still *possible,* is it not?

However, the question of whether a single molecule ever came on stage is still rather irrelevant. To understand the true nature of the problem, study the following scenarios:

- *Assume* that the DNA sequence once were produced in the primordial sea. However, there are no mechanisms to read to code, and in a matter of days in the harsh conditions of the early earth the molecule breaks apart. Alone the single molecule is completely void: There is information for the structure but the structures never become actually constructed.

- *Assume* that there is a complete body of an animal being washed on the shore of the primordial sea. However, if it is dead it is dead and that is it, it never comes alive again. Now there are all the structures ready, but the functions have forever ceased; the dynamic attractors cannot be instantiated by any amount of energy.

- Finally, *assume* that some living Robinson lands on the shore of that primordial sea. However, there are no *other* live forms already there, and there is no food to eat. Even though there are all the necessary structures and processes appropriately running in the body, there will be death in a matter of weeks.

Indeed, it is evident that there cannot exist life alone, isolated from other life, or from its natural inhabitat. Life is manifested in interaction with the environment — or, when putting this in more pointed way, life (or ability to host life) is the *property of the environment.* The problem of life is not about explaining a single molecule; the whole ecosystem should be explained simultaneously. It is evident that this observation makes the problem even more difficult. True, the one fixed molecule would be simpler to explain than a dynamic system with active interactions — but this is what life is. The role of the single molecule in explaining life is like the role of a single logic formula when explaining intelligence.

When seen in the correct perspective, there are no paradoxes here — it is only the traditional ways of thinking that are paradoxical. Concerning the evolutionary processes there also exist strange intellectual dead ends: For example, it is often assumed that the evolutionary fitness determined by the ability to produce offspring most efficiently, or by the ability to adapt to new environments most efficiently. However, the simpler an organism is, the faster it is to reproduce and to modify itself — leading to degeneracy of structures, and world power of bacteria! What is more, the assumed power of the mechanisms of natural selection also seem to be a myth: Those who have done random search in a high-dimensional space, know how notoriously inefficient it can be.

Alongside with artificial intelligence, there are efforts to construct *artificial life* (for example, see [28]). However, these efforts are plagued by the same problems as the behavior-based AI research: It seems that the emphasis is on superficial patterns. The criterion of relevance is based on how interesting a simulation looks. And it is all computer programs and algorithmic procedures — what you program there is the only what you get out; only what you can think of, you can implement. But it seems that life is an emergent phenomenon so that its essence cannot be captured in definitions. Trivialization of the complex questions only results in what one could call (following the AI terminology) "shallow life".

The new view of cybernetic systems as pursuing balance is fundamentally different from traditional intuitions. It has been assumed that "interesting" complex systems are "at the edge of chaos". When studying the processes of life, the mainstream view is expressed by Ilya Prigogine: Life is "as far as possible" from balance, whereas death means final balance. Erwin Schrödinger phrased this as "What an organism feeds upon is negative entropy; it continues to suck orderliness from its environment". Also in cybernetic systems, static balance means death — but a living system is characterized by *(thermo)dynamic,* non-static balances. The ways of thinking have to be inverted: Whereas a living thing is traditionally assumed to play an active role, now it just has to adapt to its environment; it is the environment that pumps disorder into the system, and life processes try to restore balance, or homeostasis. The system controls the environment, yes, but it is the environment that dictates how it is to be controlled.

It seems that the neocybernetic starting point is useful, capturing the correct intuitions here: The goal of a system is the ability to find the best balance with the environment. To implement this, there are not only structural adaptations available, but continuous matching processes take place to fine-tune the structures. And in evolution, becoming structurally more complex is the method to reach better match with the complex environment. These issues are studied closer in what follows.

9.2.2 Balanced autocatalysis

The definitions of what life is are very intuitive, and no matter which set of characterizations is selected, there always exist counterexamples. In the neocybernetic perspective, the following definition is employed here: *Life is higher-order balance with the environment.* What kind of assumptions are needed to make this definition plausible?

Looking from the point of view of the end result, seeing the living organism as being at the mercy of the environment makes it seem very volatile. However, analyses must be started in the bottom-up direction: Starting from the simplest of environments and proceeding towards more sophisticated ones, always making the outer system supervise and control the inner one towards local heat death, puts the system into an active role. Applying the vision of inversion of the arrow of entropy, the problems seem to become solved one by one, balances being restored in each phase separately. Getting to the higher levels, bigger picture is seen, the entity becoming better and better controlled, keeping the emergence of sophistication in the developing system thermodynamically consistent. The assumed balance with

So, start from the bottom and select a narrow view of the environment, let the system adapt there, and only after that widen the view. The transition from intuitively non-living structures to living ones becomes smoother; on the bottom of hierarchies one has *chemical evolution.*

One of the central prerequisites for life is the capability of reproduction. The simplest example of chemical reproduction is demonstrated in *autocatalysis,* where a chemical catalyses a reaction where this same chemical is produced; the autocatalyst thus can make copies of itself. Assume that for some chemicals A, B, and C there holds

$$\begin{array}{rcl} \mathrm{A} & \Leftrightarrow & 2\mathrm{A} \\ \mathrm{A}+\mathrm{B} & \Leftrightarrow & \mathrm{C}. \end{array} \tag{9.2}$$

The autocatalyst A acts like switch, activating the reaction producing chemical C from B. If there is chemical A present in the system to begin with, it will forever continue to be there no matter how much the solution is diluted. The chemical A thus characterizes the functioning of the system, selecting functions by making the corresponding reactions possible that otherwise would never take place. In practice, the autocatalytic reaction chains must be more complicated than the above one; it has been shown that the probability that a random set of chemicals is autocatalytic becomes high under certain assumptions.

Autocatalysis makes it possible to explain inheritance of functions between chemical systems. Indeed, autocatalytic sets are seen as the explanation for origin of life by Stuart Kauffman and others [44]. However, there are theoretical problems: Looking the chemical reactions syntactically, as a cookbook, there seems to be an explosion of chemicals. There is no self-regulation in the system, and there seems to be no emergence of structure. Indeed, it seems that autocatalytic systems typically only produce sticky tars, ending in deadlocks. In the neocybernetic framework this problem is solved: The system consists of balance reactions that proceed only in favorable conditions. It is the environment — or the reaction set itself — that takes care of self-regulation, the autocatalysts determining the spectrum of possible degrees of freedom in the chemical system. And the *function* is more relevant than the *structure.* In a system of autocatalysts function is manifested without solid form; structure is of secondary importance. But as illustrated in the next section, the physical properties of the world can make structures automatically emerge without explicit maintenance.

9.2.3 Chemical evolution

To understand the life processes in their simplest form, it is here assumed that chemicals participate in equilibrium reactions, as presented in chapter 1. To facilitate the emergence of something more interesting, three basic hypotheses concerning the reactions are made:

1. There are autocatalytic chemicals present among chemicals.
2. There is a medium available where interactions can take place.
3. There are mechanisms available for keeping chemicals together.

In the simplest case, this means that there is liquid water for chemical solutions to react in. Using the traditional vocabulary, one can speak of *primordial soup,* where there are chemicals and energy available (for example, see [21] and [88]). To keep chemicals together, it can be assumed that there are some kind of partially isolated "droplets" in the medium (see Fig. 9.6). The physical environment makes the droplets behave like "proto-cells". The growth of such droplets can be explained in terms of good match between the environment and the autocatalytic set characterizing the contents of the droplet: The reactions are active, keeping up the "metabolics". Chemical properties determine the internal balance in the droplet, but it is physical phenomena like osmosis and surface tension that together determine the size of the droplet, and whether it splits up. Because of the geometric constraints, the chemical reactions in the droplet are also affected: As there is less surface, there is smaller total intake of chemicals; and if some of the droplets is surrounded by other droplets, it experiences a very different environment, thus perhaps exhibiting different reactions, and different chemical functions.

The droplet has to maintain its integrity, so that it does not dissolve in the surrounding water. This can be assured if the contents of the droplet are, for example, based on fatty acids or some gels. It is the chemical reactions within the cell that have to provide also for the supply of this substrate. In more sophisticated cases the proto-cell can have some membranes that are based on phospholipids or other compounds with *amphipathic* character, having hydrophilic and hydrophobic parts. The more complex scenarios can employ the ideas of *vesicles, globules,* or *micelles* to host the reactions, having restricted exchange of chemicals and energy with the environment. Similar scenarios have been proposed a lot in the literature — but even though structures that resemble "cells" can emerge in a rather autonomous manner, it is clearly not such physical structures only that characterize living systems.

The key difference here as compared to the standard autocatalysis models is that the proto-cells are not whatever droplets, but they can host complicated sets of balance reactions. Such an equilibrium system is a *local mill of entropy.*

The entropy considerations in the beginning of this chapter were rather abstract, and can be applied only as evolutionary processes are seen from outside. But what are the mechanisms — how is the increase in entropy manifested in the lowest level of a developing proto-cell? A well-balanced proto-cell offers a good platform for biologically relevant functionalities to emerge. The neocybernetic structure has evolutionary advantage as seen in the thermodynamic perspective.

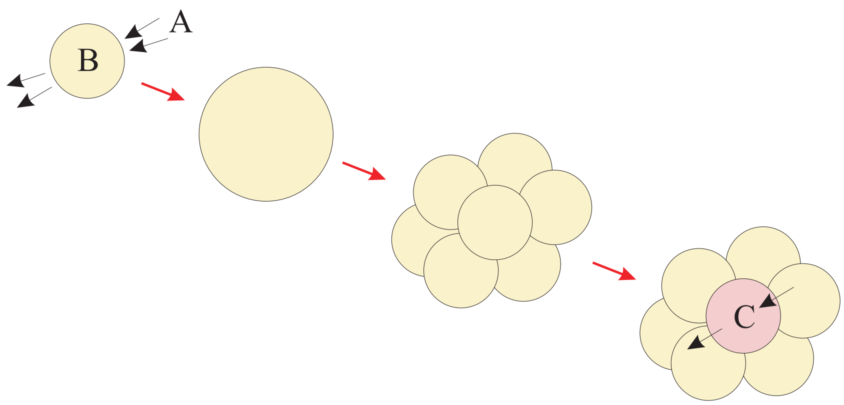


Figure 9.6: Illustration of how there can exist "cells" also without actual hereditary material: There can be nutrient intake, cell growth, reproduction, and even differentiation in the low level with no explicit developmental control. **a.** Droplet containing a set of autocatalysts is characterized by a reaction, say, A→B. **b.** There is plenty of chemical A available, so that concentration of B increases, and osmosis makes the droplet absorb water and grow in size. **c.** The droplet splits up in smaller ones because of the weakening of the surface tension; adhesion keeps the droplets still together. **d.** The middle droplets experiencing a new environment, different chemical reactions become appropriate, reaction now being B→C; such developments continue depending of the mixture of autocatalysts

First, the "heat death" within the system makes it possible for the very fragile molecules, like proteins, to remain whole, balance thus promoting survival; second, the balance makes it possible for very vague phenomena to become magnified — the "signal-to-noise ratio" in the system becomes high, revealing the remaining information in the signals, balance thus promoting more consistent evolution by making the appropriate adaptation direction better visible. In a well-maintained balance, it is enough that there is some special chemical that can only be utilized by the new cell type with a special set of autocatalysts; this gives it the adequate competitive advantage. Only if there exist alternative solutions (in terms of chemical solutions), secondary aspects like reproduction speeds become relevant in competition. Thus, around the balance, increased complexity is an evolutionary advantage, and proto-cells hosting more complex (and typically slower) reactions will survive in the chemical evolution.

No specialized "cell organelles" are needed to implement the basic cell-like functionalities; it is just assumed that the proto-cells are not completely isolated but search for the balance of reactions in the prevailing environment. It is easy to imagine what can happen next: Different proto-cells or cell groups can start exhausting each other's surplus products, and become *symbiotic.* From the point of view of a single functional unit, other ones start to do the "sanitation", exploiting the excess "waste products" — otherwise the reactions cease, the proto-cell suffocates, or, at least, its well-being becomes jeopardized. But, as seen from the point of view of the neighbors, such excess products are available resources, deviations from the nominal to be exploited. If the different kinds of proto-cells are dependent of each other, they probably grow and divide at the same rate, following the cybernetic balance; this is a rather plausible route to "multiglobular" systems (again, see 9.6). Cells without partners starve and become outnumbered. As seen from outside, different cell groups represent different sets of reactions, so that functional differentiation starts taking place. Control of the flourishing diversity is local, following the cybernetic principles; because of the physical realm, the experienced environments differ, and differentiation among cells start taking place.

When looking at the early development of an animal embryo, this kind of scenarios of differentiation is exactly what seems to take place even today in a fertilized egg: First the prototype cells form a *morula* or *blastula,* where the totipotent cells start differentiating depending on their environments. In a way, the Haeckel's intuition (see 7.3) should apply also to the earliest phases of life, giving motivation to the above studies: Individuals repeat the whole sequence of becoming alive. Perhaps there is something to learn in today's developmental processes when trying to reconstruct the origin of life. Of course, there still is an essential difference: Even in the simplest cell today there are the instructions, or the DNA code readily available, and the development is in this sense preprogrammed. But in the stem cell phase of morulas, no genetic imprinting has yet taken place. The simplest processes need no instructions; genetic code is only needed when there are alternative routes to select in the development. Perhaps the genes only started orchestrating the natural processes.

9.3 Codes and beyond

Can the above-like non-genetic, strictly chemical behaviors be called life? Today's life forms in biosphere are all characterized by genetic code, and it seems that there is a huge leap from non-genetic to genetically controlled. However, it seems that evolution towards such more sophisticated control of structures can still be explained in a rather consistent way, and no giant leaps are needed.

9.3.1 Towards programmed structures

The functionalities in the proto-cells need not be something clever or preplanned, as long as they exploit the chemicals available in their environments and produce something else — in short, being successful is capability of being active, exploiting the available chemical resources. The environment is not predestinated, as it is the surrounding set of successful proto-cells that *create* this environment. When there is appropriate accommodation, the system as a whole starts looking "clever" — but only as seen in retrospect.

When studying the possibilities of more complicated functionalities to emerge, one needs to distinguish between two separate things: First, there is the ability to reproduce, and, second, there is the ability to modify cellular metabolics. Traditionally, it is assumed that it is the same solution (genetic code) that is responsible for both of these capabilities — but this need not be the case. It is the autocatalysts that have the reproduction capability; some other chemicals can be multifunctional ones. In the lowest level, it is enough that some chemical operates in different ways in different chemical environments. For example, inert and active states can be toggled depending on the environmental conditions like pH or temperature. This means that the reactions are nonlinear. The operating modes of the cell being integrated in the chemicals themselves, the cell functionalities are accordingly changed when the environment changes.

In the proto-cells, genetic code is also not necessarily needed to control behaviors; not even any complex molecules like nucleic acids or amino acids are needed in the beginning. No code reading capability is necessary to begin with. Of course, it is practical if the two presented capabilities, reproduction and multifunctionality, are combined in a single autocatalytic molecule. And — as seen in retrospect — it seems that DNA has outperformed all other mechanisms. The combination of DNA as the code and proteins as the tools for implementing functionalities is a very versatile combination, offering almost as flexible platform for different kinds of chemical structures as the neural machinery offers for cognitive structures.

Still, all information that is inherited needs not be transferred in the form of DNA. The *Lamarckian* theories have been neglected because it has been claimed that there are no necessary mechanisms to implement such views — however, also in the highly developed forms of life, there *are* other mechanisms available. It need not be assumed that the initial state of the stem cells is completely null; there can be some chemicals that follow the genetic material into the gametes, being manifested in the tsygote. This kind of inheritance can be called *epigenetic,* being also related to *genetic imprinting.* However, it is not any acquired properties that can be inherited this way; it is the commands of which of the available genes are activated in the beginning. Another issue is that it has been recognized that the microbial symbiotic fauna seems to be also inherited from the mother. As has been recognized, this symbiotic inheritance can essentially affect the metabolic processes that are activated in off-spring.

Such symbiotic systems illustrate that everything needs not be coded in the same genome in a centralized manner. Coordinated operation and reproduction is possible without sharing any genetic material. For example, *mitochondria* in cells have their own genetic codes; *lichens* are associations of a fungus with a photosynthetic partner that can produce food for the lichen from sunlight. All the subsystems are still based on DNA of their own; these codes need to have coevolved.

Even though there were only a single set of codes, there is need for coevolution. For example, the trinity of DNA, RNA, and proteins necessarily had to be there from the very beginning in some simple form, even though the roles of the components need not have been so clear-cut — the theory of the "RNA world" as the immediate predecessor of the modern life forms probably cannot hold. As discussed in the following section, the developments from the beginning of life to the present day have to be smooth — the basic structures cannot be changed abruptly. Even though there are tensions towards more sophisticated structures, it would be difficult to understand huge sudden steps in developments. The interesting question that remains is what the actual autocatalytic set of simplest proto-DNA, proto-mRNA, and "protoin" is.

Continuity (and differentiability) of functions is the key to efficient optimization in mathematics; otherwise, one only can do random search, and in a high-dimensional space this is extremely inefficient. How can continuity and consistent adaptation be reached when the functions are based on discrete genes? And, further — how can the very discrete nature of structures in the phenotype be explained if the functions are continuous?

The key point is that it is not a static one-to-one mapping from the genotype to the phenotype, but it is dynamic processes that implement the mapping, the static-looking patterns being the final dynamic equilibria. As the genetic system is in contact with its environment, it searches the balance; redundancy of the genes, and the quantitative nature of gene expression makes it possible to reach continuity. And because of the sparse-coded nonlinear nature of the genes, there can exist various equilibria: Minor changes in environmental conditions can result in very different outcomes, giving raise to emergent structures.

Genes are modified in a Darwinian process of mutation and crossover; however, the genes are not actually optimized. The main role of evolutionary processes is to generate variation: The goal is to supply material, a pool of alternatives, whereas the local balances within a cell finally select the appropriate genes, revealing the actual potential and limits of the new genetic combination. The genetic process determines the (sparse coded) subspace in the metabolic space, and other processes are utilized for final optimization within those subspaces. The genes only determine the *potential* in terms of available degrees of freedom, whereas the environment determines the *actual,* the location of the equilibrium in the search space. In each cell the same functionalities in latent form wait to be activated. Optimality of solutions is defined in a very local and immediate fashion, there is no need to wait feedback from explicit "goodness" evaluator, with the delay being of the order of one generation — level of match with the surroundings suffices. Yet another fact needs to be recognized: There is no global single fitness criterion. Each variable is being matched more or less independently, so that, in a sense, "parallel processing" for fitting the data is implemented, further enhancing the adaptation speed.

The genes are hierarchic, and there is often accumulation of various individual genes that is needed to implement some more sophisticated functionalities. The benefits of the genes are visible only after the whole structure is completed — how can the sudden emergence of such complete functionalities be explained? However, the local minima are not necessarily very far apart, and the chain of gene activations can still be reasonably cut in subparts, as studied below.

9.3.2 Case: Development of an *eye*

It has been claimed that evolution theory cannot hold — there exist a plenty of highly complex structures that are functional only when they are complete. As long as the structure is not yet fully developed, the infrastructure for it is only a burden, and evolutionarily disadvantageous; this should mean that the barrier between the local fitness maxima is too wide to be crossed. The complex organ should have emerged immediately, without intermediate steps, and this is simply too improbable. A typical example is the *eye* — an example that has been studied widely in literature.

However, it turns out that there is a path from no-eye to a complete eye consisting of simple gradual steps where each stage is evolutionarily beneficial.

Indeed, it has been observed that the eye has developed various times in different branches of the "tree of life", and the solutions are not unique. Below, one simple scenario is presented.

Still, it seems that the population-level feedback loop between better properties and consequently more probable survival is too inefficient to support the consistent development of structures — and, especially, the simultaneous development of separate functionalities seems like a too lucky coincidence. For example, enhancements in the eye cannot be exploited if the processing of the neural signals is neglected; the eye and the brain have to develop in a somehow orchestrated fashion. — Indeed, it seems that neocybernetics may offer some tools to understand such dilemmas, as in that framework the *fitness criterion can be distributed.* All cells simply try to maximize the emergy they receive from their environments, whatever form this variation takes; emergy is not only physical nourishment but information in general. There is no need for external evaluation; variation or information can be regarded as beneficial no matter whether that information can be exploited locally or not. As the system itself is quite well-balanced, increasing variation reflects enhanced coupling with the environment. Increasing excitation in the eye means increase in nerve cells, and increasing excitation in the nerve cells means motivation for brains to develop. It may be (it is) so that at the higher (proto)animal level better processing of the visual signals means better possibilities of responding to the threats and opportunities in the environment, thus improving the survival of the organism as a whole, but the underlying organs and tissues can see the developmental gradients more instantly.

In Fig. 9.7, a simple scheme is depicted where the development of an eye can be understood. Each step in the series of stages means more accurate detection of behaviors in the environment — first, there is the capability of seeing whether it is dark or not; second, the direction of the light source can be detected; after that, it becomes possible to tell patterns from each other, with ever increasing accuracy and sensitivity. The presented development process is by no means unambiguous — for example, compare it to the compound eye structure of a house fly.

As compared to discussions in chapter 8, it is again interesting to study the relationship between the genotype and the resulting phenotype — what the "interpreter" is like, and how "semantics" in the physical and physiological environment can be defined in the above hypothetical case. Evidently, the results of interpretation are this time determined by the limited information delivery in the physical domain. The physical dimensions are crucial — it can be assumed that the signals are transferred in terms of chemicals; as there is no way to control the spreading of the chemicals, the environment of cells is determined in terms of *concentration gradients.* This means that the codes have to be essentially location-based, the fixed points in the configuration of cells being determined by some activated cells expressing some special genes and producing some signaling chemicals: It can be assumed that the concentrations of the diffusing chemicals decay monotonically (exponentially?) as the distance to the imprinted cells increases.

Chemicals, or, indeed proteins, are the only immediate outcome of gene expression — how are the chemical concentrations changed to physical properties of

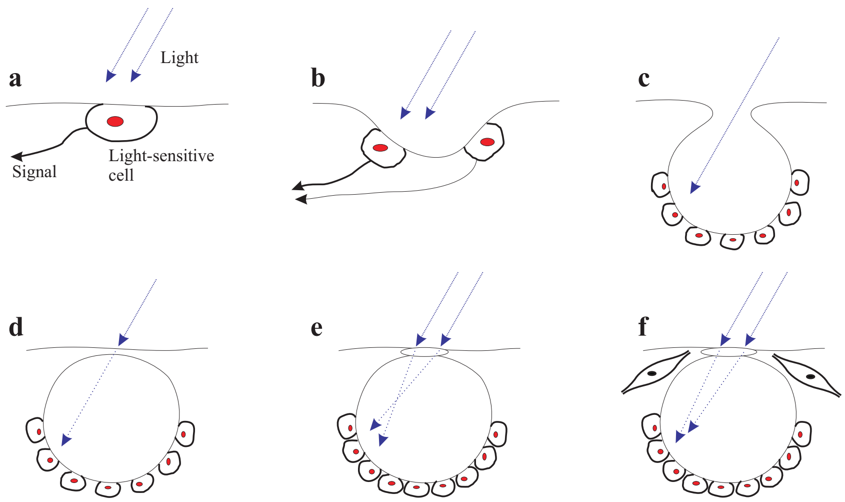


Figure 9.7: The development of an eye can be explained as a continuous process of enhanced information retrieval. In **a**, there is just a single light-sensitive cell — but knowledge of whether it is dark or light is already valuable information. In **b**, physical reasons can make the cells differentiate, as the amount of absorbed light depends on the direction where the light is coming from — and knowing *where* there is light can be crucial information. In **c** this differentiation has proceeded, so that what one has is a simple "needle's eye" camera — it is already possible to distinguish between different light sources. Later, in **d**, the process of increasing the eye resolution ends in the opposite walls of the cavity merging together — this, too, is beneficial as the robustness of the proto-eye increases, the sensor cells becoming isolated from the environment. It is clear that there is evolutionary advantage if this filter layer becomes more transparent — and as it does, in **e**, one has a lens, making it possible to reach much higher light sensitivity and resolution at the same time. Finally, in **f**, the ready-to-use muscle functionality is employed to deform the lens, thus increasing the adaptation capability of the eye. It is clear that the brain has to co-evolve to make use of the available new information — if the information is not cleverly exploited, the evolutionary pressures vanish and the consistent developments cease

tissues? Enzymes and transcription factors can promote cell metabolism and reproduction, and possibility of increasing activity means increasing biomass. The thickness of tissues is related to numbers of cells; further, the transparency of cell layers is related to thickness, etc. — when the physical properties of cells become manifested, all physical functionalities are available that are prerequisites for imaging and image processing. Again, the physical constraints result in emergence of smart-looking structures — as seen in retrospect: Development of cavities facilitates differentiation among the light-sensitive cells, etc.

According to the above lines of thought, below is a simplified example of what the "eye program" could look like, when a "transcription" from DNA into an explicit pseudocode is carried out:

```
1.  IF ‘‘location’’ ≈ p1
    THEN imprint ‘‘EYE’’:
            utilize light sensitivity, emit ‘‘eye’’ and NGF

2.  IF ‘‘SKIN’’ AND ‘‘eye’’ ≈ p2 AND ‘‘lens’’ < eps
    THEN split up

3.  IF ‘‘location’’ ≈ p1 AND ‘‘EYE’’ ≈ p3 AND ‘‘lens’’ < p4
    THEN imprint ‘‘LENS’’: develop transparency, emit MGF

4.  IF ‘‘lens’’ ≈ p5 AND ‘‘muscle’’ < p6
    THEN imprint ‘‘MUSCLE’’: grow towards MGF, emit ‘‘muscle’’
            etc.
```

In the above code, each of the four rules represents a function of its own, or a gene (or set of redundant genes), as listed in order of assumed activation. The first row in the rules describes the control part, the rest determines the “actions”. It needs to be recognized that the “interpreter” for the code is distributed, running separately in each cell; the code is identical in the cells, so that differing reactions are caused by the local environments. The communication and coordination among the cells is implemented trough special signaling chemicals that are in the code denoted by quoted lowercase names. The quoted uppercase names denote imprinting — that is, the cell is assumed to have reached some specific state and its role has been determined.

The first gene becomes activated if one (or more) chemical levels match the preset value(s) p_1; such “location” signals truly exist in a real embryo where, for example, they implement the anterior–posterior an dorsal–ventral asymmetries. The comparison operation is streamlined here — this genetic control perhaps has to be composed of several elementary toggles as presented in 6.2.4. When this gene is activated, the corresponding cell will forever have the role of an eye, its special property being light sensitivity. Simultaneously, it produces chemicals: when the signal “eye” diffuses outside the cell, the neighbors can detect its existence, and nerve growth factor NGF starts persuading nerve cells to connect to the cell. The second gene can only be activated in a skin cell: If there is an appropriate distance to the eye cells (chemical “eye” in the surrounding tissue having decayed to the level p_2) and there is no “lens” signal present, the cell starts reproducing, thus making the number of cells grow, causing the skin get wrinkled, leaving the eye cells on the bottom of a cavity. It is assumed that neighboring cells automatically attach to each other when they are in contact. The reproduction ceases as the lens develops — this happens in the cells that are in the center of the eye area but are not eye cells. In evolution, such lens cells develop towards better transparency. The lens cells also secrete nerve growth factor, trying to persuade muscle cells to attach to the lens. The last rule represents the muscle as being exploited by the eye — however, muscles are of course general purpose structures, and they can be activated also through other sequences.

The "muscle" example above is characteristic to genetic systems — once the muscle functionality has been "invented", it is readily available and can be exploited in different organs. It can pop up in different structures — and if it is beneficial (as it is in the eye), this new functionality is supported by later evolution. The genetic substructures are ready to pop up as soon as such functionality is needed; in this sense "genetic design" is like *functional programming.*

The functional structure of the genes as proposed above is very simple offering just the coarse framework for physiological structures; within this discrete representation, there is continuity in the structures. Quantitative fine-tuning is possible in terms of the parameters p_1, p_2, p_3, p_4, p_5, and p_6 affecting the eye dimensions. As the two halves of the genome are inherited from differing individuals, the threshold values in the parents typically become averaged in the offspring. This is an efficient optimization scheme when there is only a narrow region of acceptable parameter values available. Different-looking species are possible when different parameter values are selected — and, indeed, as most of the genes are common to all life forms, some kind of fine-tuning of their effects is necessary.

9.3.3 Optimality in mechanical structures

Not everything can be coded in genes: After all, genes represent central control, and if employing only them, behaviors would not be tuned maximally — or, at least, adaptation towards the optimum would be very slow. The genetic machinery needs to be accompanied with better adjustable mechanisms to reach the fine tuning. In a way, it is as it is with cognitive systems: The lowest level (chemical concentrations or synaptic weights) is continuous, the "intermediate level" (organ structures or conscious thinking) is discontinuous, and the highest level (complete optimized organism or automated behaviors) is again more or less continuous and optimizable; to facilitate real-life survival, automation of slow cognitive processes has to take place, and, similarly, the final polishing of the structures in living bodies takes place after the actual implementation of the codes. Especially, this means that evolution of fitness cannot be based on so delayed mechanisms as it is assumed when speaking of natural selection; more immediate feedback mechanisms are necessary.

In principle, the neocybernetic optimization principles can be applied in any environment and at any level. However, the intuitions about uniformity among signals collapse when the system is a sophisticated functional entity; the vision of a cybernetic system as reflecting its environment also becomes far-fetched. The relationship between the system and its environment becomes blurred as it is the other organs that deliver the input signals to an organ — this environment also changes, and one should implement optimization for all systems simultaneously. Can the basic neocybernetic model of separate system and environment be applied any more when all signal are internal ones? This issue can be studied when looking at Fig. 9.5 again: When the system and the environment become one, the only thing that remains outside is the feature extraction. Now this generation of features, or manipulation of measurements, is not artificially constructed by some designer, but it reflects the effect of how the real world distorts the signal transmission process among the organs.

To make the above discussion more concrete, let us concentrate on the question in what sense the *outlook* of an organism can be captured in compact formalisms and optimized therewith. To proceed, one needs to imagine how the organs act as probes deforming the "steel plate" around them (see chapter 3). The outside world is not known, but iterative adaptation within the individual organs (or cells) still optimizes the system.

To have some more background, it is necessary to get acquainted with the techniques of modeling mechanical systems. In *Lagrangian mechanics* it is observed that the Newtonian laws of motion can be reformulated as optimization problems: Along the motion trajectory, the time integral of the quantity $L = W_{\mathrm{kin}} W_{\mathrm{pot}}$ reaches its minimum value, where W_{kin} is the kinetic energy of the system, and W_{pot} is the potential energy. Applying the vector of generalized coordinates q, the kinetic total energy can be expressed as

$$W_{\mathrm{kin}} = \frac{1}{2}\dot{q}^T I \; \dot{q}, \tag{9.3}$$

where I is the *inertia matrix,* and the vector$\dot{q}$ stands for the generalized velocities (translational or rotational). Does this not look familiar? Indeed, when defining $x = \dot{q}$, the basic neocybernetic cost criterion can be interpreted in this framework:

$$J = \frac{1}{2}x^T \mathrm{E}\left\{\bar{x}\bar{x}^T\right\} x - x^T \mathrm{E}\left\{\bar{x}F^T\right\} F. \tag{9.4}$$

Now, $\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}$ can be interpreted as the inertia matrix. If F is the vector of forces and torques that can sustain the corresponding velocities, $\mathrm{E}\left\{\bar{x}F^T\right\}$ becomes some kind of a *viscosity matrix* and the latter term in (9.4) is the viscous work (or power lost in movement). This is an extension of the Lagrangian thinking: Forces in the assumed system are *non-conservative,* as the "potential" is not free of the velocity variable.

It is an open question whether the above cost criterion truly has relevance in real life. Yet, if it does, this cost criterion offers a framework for analysis of natural life forms; what is more, it makes it possible to create "cybernetic designs" in life-like (biomimetic) structures. It is clear that the balanced designs are optimal in such a sense that *maximum amount of correlated forces are transferred into movement.*

Assume that muscle cells (forces) and sensory neural cells (velocity measurements) are modeled together. The relationship between these variables is determined mainly by the limb configuration — this relationship implements the "feature generation". When the integrated system becomes optimized in the neocybernetic sense, constructing a statistically balanced model for the relationships between the variables, the system speed and agility become optimized automatically: There will be maximum possible velocity in structured (sparse coded) directions with the minimum effort. The iterative optimization process can end in outlooks that differ very much from the initial. Local adaptations in the structure (as accumulated in the inertia and viscosity matrices) are reflected in the increasing overall "fitness" of the global structure. Even though the goals of adaptation are not fixed beforehand, the direction of "better performance" is known by the local actors, and the post-genetic developments are not random — there are gradient directions visible. For example, rehearsing of muscles makes it possible to learn the model between the forces and corresponding limb velocities; adaptation of this model results in ever more optimal and economical (and thus more "beautiful") trajectories — the final outlook of the body need not be coded in the genes.

9.4 Are we alone?

As discussed above, it can be assumed that life unavoidably emerges if the conditions are favorable, and if there is enough variation in the conditions so that the modeling task is non-trivial. And as the arrow of entropy is inverted, it is not difficult to imagine that intelligent life is just the next step in the inevitable development of life forms. But if the origin of life and intelligence can be explained in such a straightforward manner, one is facing yet another paradox. When there assumedly exist millions of planets that can host life, and as the evolution sooner or later results in intelligence emerging among the life forms, there is a question that was originally coined by Enrico Fermi: "Where are they?". Why cannot we see the activity of the other civilizations? There simply must be other (more) intelligent civilizations in the universe in addition to us.

When the radio frequency spectrum has been scanned, nothing "intelligent-looking" has been found, only noise has been detected in the signals coming from the stars. But there is a simple (partial) explanation available here. One can only search for redundancies in the signals — but, from the point of view of transmission efficiency, redundancy necessarily means unoptimality. A message with all redundancy ripped off looks like noise if the decoding scheme is not known. It is a very short period in the history of a civilization that transmissions are not optimized and packed; in our case it is something like 100 years only. — But it even seems that the other civilizations actively try to keep the distance, why is that? One can make some hypotheses here.

Why are they not trying to contact us — as we do, sending easily decodable signals to us on purpose? But, on the other hand, why should a civilization make a big number of itself? Only civilizations being in the early stages of their intellectual development make a big fuzz of themselves — the older and more mature ones observing this blustering sympathetically. If a civilization is to survive the turmoil periods there are in the development, the periods of chauvinistic arrogance need to be overcome. Indeed, knowing that there are civilizations millions of years ahead of us we should perhaps be a bit ashamed. It is plausible that we could not even recognize the systems far ahead of us: After the chemospheres and biospheres, our frontier systems today reside in infosphere. But after the principles of intelligence (or infosphere cybernetics, really) are fully implemented in the computer, developments in infosystems become very fast — so fast that when the time axes in information modeling processes collapse into singularities, qualitatively yet another level is perhaps reached. It is impossible for us to understand the higher-level systems: It is like biological systems facing cognitive systems — trying to "understand" the structures on the emergent level is an intellectual contradiction.

But why is this passivity so categorical — why do none of the higher-level civilizations even play with ourselves? Indeed, this consistency promises that we are on the track of something big here.

It seems that this complete silence is purposeful: They do not want to disturb us, they want to see how we manage on our own. Preservation of life and natural diversity is seen important by all intelligent civilizations — and there is a reason for that. The analyst does not want to disturb the processes he/she/it is observing. The problem of life seems to remain an eternal challenge for intelligent minds: Understanding the mystery of life is understanding survival, and it seems that it is cybernetic-like modeling over the spectrum of possible forms of life that will continue as long as the civilization lives. More material, more fresh data is needed to map different local solutions in different environments.

But it is not only curiosity that drives such galactic research — such research is necessary to maintain sustainable development. Also extraterrestrial life is facing the limits of its home planet, and the only way forward is available in the infinite space. All evolving civilizations have to be based on science and information pursuit, searching for new frontiers. The link between science and the society is the question of life and death to a civilization, even more than what we can understand today. The more intelligent a species is the more it is dependent of scientific research, as acquiring new information seems to be the survival strategy. The intelligent species necessarily have found the principles of cybernetics, and they must understand that the only way to avoid cybernetic stagnation and catastrophes is to receive ever new information, and new sources of information. It might be so that it is us as a peculiar example of living systems that provide a piece of this crucial information. What is the nature of this information, then? Of course, such knowledge is beyond our capabilities of understanding. We cannot yet see "higher-order life" in such a wider perspective where our world would just be a single sample case.

From the point of view of the higher-level intelligence, we are running just another experiment. As Douglas Adams observed in his "Hitchhiker's Guide to the Galaxy": The Earth is a giant simulator. Perhaps some day — if we pass the test of survival — we receive an invitation to the "Galactic Board of Intelligent Species" where the universal experiment designs are coordinated. And perhaps it is our maturity test — being faced by any developing civilization — to understand the cybernetic processes, tame them, and avoid the downfalls, making the succession of ever more deadly catastrophes a steady process of sustainable development.

The risk of mankind committing an explicit suicide is today a well-understood risk — but it seems that there also exist more latent threats to a developing civilization. As studied in [22], the developments in Tasmania may give us a hint of this risk:

> Some 10000 years ago Tasmania was cut off from the Australian mainland, and about 4000 Aboriginal Australians remained totally isolated. When Europeans discovered Tasmania in the 17th century, it was technologically the simplest, most primitive human society. Native Tasmanians could not light a fire from scratch, they did not have bone tools, they did not have multi-piece stone tools, they

> did not have axes with handles, they did not have spear-throwers, they did not have boomerangs, and they did not even know how to fish. Incredibly, archeological investigations have shown that during those 10000 years of isolation, the Tasmanians actually *lost* some technologies that they had carried from the mainland to Tasmania. What caused this decay in civilization?

There were no catastrophes in the Tasmanian culture — before Europeans, there were no rivals and no external disturbances to shake the system that had reached the stagnation. The smaller the population is, the faster it seems to reach the stasis — but, similarly, a planet-wide monoculture can decay; what is the difference between evolution and devolution?

Living systems seem to share the property of vitality — there is some kind of *arrogance,* tendency to grow and conquer. Growth has to be eternal, but this growth need not be physical, it can also be mental. The internal spirit is a matter of life and death — and even though the cybernetic principles are universal, the evolutionary processes cease if this spirit is missing. It seems that vitality must be explicitly maintained. What does such loss of memetic vital force look like in concrete terms? It can be claimed that it is *pessimism* — as seen in the scale of the whole civilization — that characterizes the end of culture: When everybody starts looking back into some lost paradise, trying to oppose the change, there will be decay. The system as a whole must keep up optimism and curiosity, looking forward even if facing the "cosmic angst" in front of the unknown future.

It has been said that if an individual human being wants to be happy throughout one's life, keeping up optimism and good humor, one should do *gardening,* letting living systems grow and seeing them prosper. Perhaps the same ideas are the key to sustainable, non-explosive developments in the wider scale, too — if an individual civilization wants to live "happily" ever after, it should do "gardening" of lower-level civilizations. And just as a good gardener protects its plants, nourishing and eliminating hazards, perhaps the "universal gardener", the "cosmic philanthropist", also protects its planets, looking after us ... perhaps the belief in personal gods (or UFO's), our higher-level protectors, is not completely unjustified?

Level 10

Models of Reality can be *Reality Itself*

Modern philosophy to a large extent consists of commentaries on ancient philosopher's writings. But this way of doing philosophy is outdated: The world has changed; we have so much new information about the world available today. Neocybernetics may offer new ways of thinking about the world around us. But this is not all: It can also challenge our ways of *thinking about thinking.* The new approaches address the very basics of scientific work.

Immanuel Kant started the "Copernican revolution" in philosophy, calling it the "prolegomena to all future metaphysics". Today, the same kind of ideas can be applied to any science; one could also speak of metabiology, metacognition, etc., when speaking of the principles below the surface observations. Perhaps neocybernetics is a step towards a "prolegomena to all future metascience" in general. This final chapter discusses some issues concerning the emerging "metascience".

10.1 Models are what there is

Models try to explain observed real-life behaviors applying some simplified theoretical framework. Also in this text, everything has been about constructing models of reality. And, traditionally, it is emphasized that *models are always false* — they simplify, they only capture some specific aspect of the real-life complexity; the essence of the real world cannot be captured. In principle, this starting point about deficiency of models applies also to neocybernetics. However, this is not the whole story.

10.1.1 Escape from the cave

Plato's "Allegory of the Cave" illustrates a dilemma that has been haunting philosophers throughout history:

> Imagine a prisoner who has been chained since childhood deep inside a cave. The only thing he can see is the cave wall, where shadows of objects outside the cave are cast. We know that the prisoner can only see a very limited view of reality — but to the prisoner, the shadows are the only reality he knows of. ...

The prisoner constructs his world view based on the very limited information — but he still thinks that this is the whole "truth". This may sound like a strange situation, but, indeed, also our observations are limited by our senses. In a way, we all are living in our personal caves (or cages): The reality cannot be observed directly, and there is no guarantee that our senses succeed in delivering a truthful image of the outer world. The transferred image is incomplete, but is it also distorted? Our conceptions of the world is necessarily subjective — how can one know that others share the same views?

There exist very extreme views: For example, René Descartes (1596–1650) observed that, after all, the only thing one can know for sure is that one *thinks,* and therefore this thinking mind must exist — "cogito, ergo sum". A yet more extreme view was coined by Gerorge Berkeley (1685–1753): There is no material substance and all things are collections of ideas or sensations, which can exist only in minds and for so long as they are perceived. This means that also we exist only in the mind of our creator God! Such considerations are rather fruitless — Immanuel Kant (1724–1804) was the pioneer when bridging *rationalism* and *empirism.* The perception is an interplay between what is "inside" and what is "outside": One only has access to the observation data, and it is filtered through the mental machinery; but the principles of this machinery are shared by all observers, and in this sense, there must be something in common in the subjective world views.

In a way, Kant is speaking of *internal models* and matching of data against them. And it is this model-oriented thinking that is the basis also in neocybernetics — the adopted model structure there is just another filter for the input data. The models that are based on the data covariation in the PCA style are advantageous as the data-based "semantics" can be characterized in a mathematically rigorous way. Only statistical phenomena can be captured, so that information on individuals is lost, but, on the other hand, the automated data analysis makes it possible to abstract over particulars. One can avoid predestinated, more or less subjective characterizations of the world; there is no need for explicit "ontologies" as the structures are determined in terms of correlation structures among data. The philosophical deadlocks are circumvented as "truth" is substituted with *relevance:* Constructs are important if they *exist,* what cannot be observed is not modeled.

It has been claimed that models are "out" — it has been always admitted models cannot capture the essence of systems, but when studying complex systems in particular, it seems that they cannot even capture the behaviors of them: In chaos theoretical models small deviations in initial conditions result in completely different outcomes. In neocybernetic models, however, stochastic variations are not significant. It is statistically relevant constructs or attractors of dynamic processes within the phenosphere that are being captured, and role of the transients fades away. As the approach is thus inverted, the emphasis being on balances, in the resulting emergent models one can capture *not only*

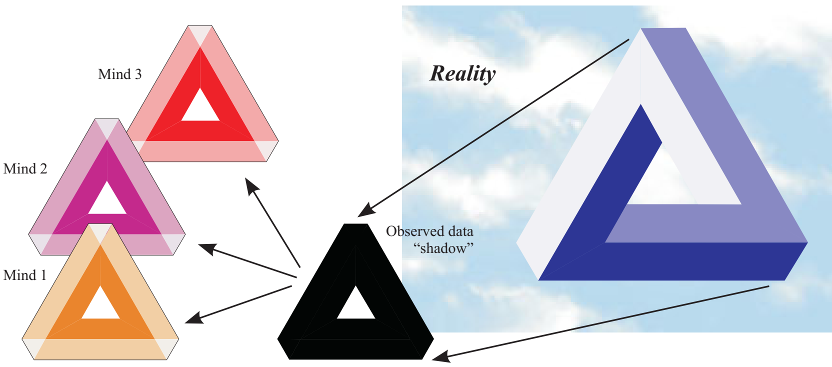


Figure 10.1: Visualization of intersubjectivity. It does not often truly matter what the reality looks like (pattern on the right), as long as the subjective reconstructions (on the left) based on the measurement data (the "shadows") are similar — the mental constructs can then be shared and are negotiable, being a solid basis for a "supermind" extending over a single brain

the behaviors but also the essence.

10.1.2 Intersubjectivity and interobjectivity

The model construction principles of cybernetic systems also applies to the cognitive domain: The mental system constitutes a "mirror image" of the environment as determined by the observations. No matter what the underlying realm truly is like beneath the observations, the mental machinery constructs a more or less unequivocal model of it; this modeling can be repeated in different minds, and the results are always essentially identical as long as the statistical structures in the observation data remain the same. This means that the separate minds then share the same mental representations, and *intersubjectivity* among minds has been reached. It may be that this mental model does not represent the real world object in the best possible way — but from the point of view of the minds understanding each other, this does not matter (see Fig. 10.1). Thus, there is a possibility of constructing yet higher level models based on the consistent world view shared by the intelligent agents.

The models in the infosphere remain intact even if the models were implemented in another medium. For example, if the same cybernetic modeling principles are copied in the computer, there will be a fundamental correspondence among the data structures as constructed by the computer, and the mental representations as constructed by the brain in the same environment. This makes it possible to reach intersubjectivity also among artificial and natural minds: The world models can be essentially the same, not only between humans but also between humans and computers. This makes it perhaps possible to reach artificial intelligence in the deep, not only in the shallow sense. Clever data processing becomes possible: The computer can carry out the data preprocessing in a complex environment, and the constructed data structures can be interpreted directly in terms of corresponding mental representations. The cybernetic computer can do real modeling, not only tuning of predetermined parameters in man-made models, applying the metainformation it then has.

But intersubjectivity can also exist in the same way directly among computers in Internet. Today's ideas of "semantic web" are plagued by the need of defining hard-coded artificial ontologies; when the network of cybernetic computers is truly semantic, computers can interact without the help from humans. When the human is dropped out of the network altogether, the possibilities of AI become practically limitless as the communication among computers and adaptations in them can take place practically instantaneously. The time axis once again contracts towards a singularity, and the emergence of yet higher-level systems can take place. In the spirit of Friedrich Nietzsche, the smart computers can host "oversystems" recognizing that their "God is dead" ... perhaps it is clever to implement Isaac Asimov's "Three Laws of Robotics" in computer hardware when it is not yet too late!

Intersubjectivity is not all there is; indeed, one can reach *interobjectivity.* If nature itself tries to construct models for eliminating free energy in the system applying model-based control, as presented before, the human trying to model these cybernetic systems can touch not only the shadows of the behavior (in the Platonian sense), but the actual essence – these models can be *fundamentally the same.* This means that if some naturally evolved cybernetic system (an ecological system, for example) is modeled by a human applying the appropriate (cybernetic) principles, this model has a deep correspondence with the system itself. Essence of systems is not in the building blocks but in the information structures. This is a very deep observation: Human can truly understand one's environment; there is a unity of knowledge and nature, and *epistemology* and *ontology* become the same (see Fig. 10.2). The model can represent the actual system losslessly regardless of the non-ideal noisy observations there are in between.

It is the common endeavor of Nature and Human to understand the Universe — and both of these are built by the Universe itself ... for doing introspection! As the reality is too complex to be modeled in the simple mathematics of neocybernetics, more sophisticated models are needed: The human carries out the task nature has given him, modeling those systems that are too complicated. In any case, the end result is the same — models are used for better understanding, for exploiting the resources and bringing them to heat death.

10.1.3 Unity of models

Usually in sciences, there is less and less in common between fields of research, as one goes deeper and deeper. In cybernetic studies, however, it seems that no matter what is the scale, going deeper and deeper makes different fields have *more and more* in common. This is due to the starting point: As it is completely distributed systems that are being studied, the analyses boil down to understanding the underlying individual agents being at the mercy of their environments. Survival there, and capability of exploiting the available resources,

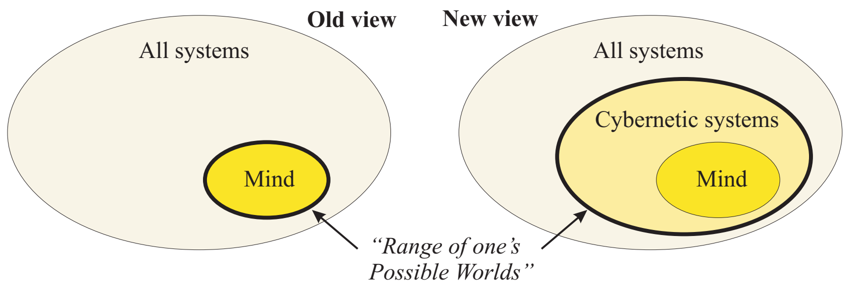


Figure 10.2: From Platonian shadows to Platonian ideals. Neocybernetic analyses can extend the range of human understanding: One can not only construct models of the systems in the environment — one can reconstruct the systems themselves in one's mind in a more fundamental sense. In the spirit of Eastern philosophies, one's world and one's mind can become *one*

can be reached when the very simple rules are followed, these rules supplying self-organization and self-regulation being always the same; when seen from above, the systems always seem to be composed of consistent balancing tensions.

Atomic physics underlies chemistry, biology is based on chemistry, psychology emerges from the biology of neurons; sociology, economics, etc., are built on the interaction of countless individual humans. But closer analyses reveal that there are also interactions between these levels, and the hierarchy of disciplines is becoming a matrix. *Consilience,* as discussed by Edward Wilson [90], is an intuitive belief that the fields of human intellect are fundamentally the same. In the neocybernetic framework this intuition can be extended: The unity of disciplines does not apply only to sciences. As there is the tight coupling between mind and matter, there is a relationship between mental and material cybernetic systems, so that everything becomes a strange, fractal but holistic mosaic. Neocybernetics makes it possible to carry out concrete analyses, as everything can be understood in terms of measurable information: This means that the domain of natural sciences is still extended — but, at the same time, the fields of philosophical speculation are extending.

What are the neocybernetic models like — some characteristics are summarized. To capture the emergent pattern, the time axis is eliminated in the models, so that the final state of dynamic balance is represented. When the static structure is found, its natural dynamics can be derived from its internal tensions, abstracting over individual trajectories, so that one has model over plausible behaviors. In the slower time scale, there is adaptation towards constant stiffness of the system as experienced by the outside observers — this characterizes the evolutionary goal of the system. As the cybernetic model is general, applicable for analysis of very different kinds of systems, analogies are a valuable tool for understanding complex systems.

When looking the model closer, it is interesting to note that at certain level of model complexity it is *causal* representations that become the most appropriate.

This claim has a very solid motivation, as studied in chapter 3: Modeling is done in terms of own actions and corresponding reactions from the environment, or x is the cause for Δu. In this sense, the hierarchy of models becomes a hierarchy of causalities.

Heinz von Foerster essentially claimed that the mind (a cybernetic system) cannot understand systems of the same level of complexity, or another cybernetic system — there is the problem of *infinite recess.* Analysis of such intertwined systems, or "Second-Order Cybernetics", ends in problems: Thinking about a system necessarily invokes both levels, first-order and second-order systems, and there is a mess, one cannot distinguish between them in analysis. In neocybernetics, there are no qualitative leaps between the levels of systems; the hierarchy of levels collapses into a singularity. Human can be liberated from the loop of analysis, there are powerful conceptual tools for artificially "understanding" the inner and outer processes alike: The human thinking is just another system to be analyzed.

10.1.4 About "truly general relativity"

Unification of models sounds like a panacea: Assumedly one only has to write this one world model once and for all? However, even though the model of the environment is objective and deterministic, as studied above, it is *still not unique,* and there is still diversity of systems.

There will never exist a complete world model as the model is relative to the observer and observations. How the *potential* becomes *actual* and in which form, is dependent of the coupling between the observer and the system — or in other terms, determination of the features that are used to characterize the system. And, further: As studied in chapter 4 in context of the 'steel plate" analogy, the environment being measured (the world) is deformed to match the probes. It is not only time (and space) that are relative — all information that is acquired is relative. The Heisenberg's "uncertainty principle" does not only apply in microscale, the same compromising has to be accepted also in macroscale, when observing large complex systems: If coupling is made tighter, the environment is deformed and the measurements change. And when doing observation, it *is* necessary to firmly push the systems to make them reveal the real structures of the underlying tensions. World consisting of elastic systems tries to yield and escape measurement, information being redirected in the non-restricted dimensions — one could speak of a "generalized Le Chatelier principle'. As Heraclitus observed, "nature loves to hide".

There is an age-old philosophical dilemma: *If a tree falls in a forest with no one to hear it, then does it make a sound?* — similarly one can ask whether a system exists if there is nobody to model it. If there is variation, it is information only if it is exploited by some system. George Berkeley claimed that "to be is to be perceived". To exist, or to be relevant, is to be in interaction with other systems and affect them. Systems and models are mixed, and experimenting and identification takes place all the time — the world is truly a holistic place. But also subparts can be studied, because such subpart reorganizes to fit the observer's expectations. A human is also a probe, an interface of a memetic system into the world.

The system is different when it is seen by some other — and the world as a whole is different when seen by some other. This observer (model constructor) need not be a human, it can also be another system. It is not because of subjectivity of models — it is because the system truly is different when seen through other eyes. If one selects some variables to construct a model with, the environment is reformed to obey the assumptions. The world view can become consistent because it is forced to become consistent, and the starting points can be motivated afterwards.

Man is the measure of all things — and this can be extended in the cybernetic spirit, as *a system is the measure of other systems.*

Speaking of relativity and numerical nature of models can lead to incorrect connotations. Statistical models sound like something uninteresting — if statistical averages and expectations are concentrated on, nuances vanish, and one only has some obscure “genderqueer models”. However, in neocybernetic modeling the emphasis is on variation and differences; sparse coding in the final models means that structures that are relevant as independent entities become separated. In the spirit of Eastern masters:

> Before Zen, men are men and mountains are mountains, but during Zen, the two are confused. After Zen, men are men and mountains are mountains again.

10.2 About “new kind of science”

Stephen Wolfram [91] prophesized that old ways of doing science are powerless when studying complex systems — traditional mathematics has to be forgotten. As has been observed, such absolute pessimism is probably not motivated; one only needs to apply new interpretations and fresh ways of thinking when doing mathematics.

10.2.1 Mathematics in a change

As has been already observed, the potential of “old science” are not yet exhausted. But the ways how observations are interpreted are changing; and because it is this observation data that is the basis of one’s world view, one is actually facing *a new kind of world.*

Wolfram’s vision of ignoring mathematics altogether is futile, as *any* reasonable formulation of logical thinking is mathematics, after all. And there exist other motivations for sticking to the old modeling tools: As Eugene Wigner [87] expressed it, “the amazing applicability of mathematics to the physical world is a mysterious, undeserved and inexplicable gift”. The main philosophical problem is not the applicability of mathematics to our descriptions of physical reality, but, rather, the major role of human-created mathematics in the *discovery* of new phenomena. What is more, it often turns out that it is the very simple mathematical machinery that only is needed — perhaps the reason for this is that in neocybernetic models, as studied before, no sophisticated mathematics is truly needed? Cybernetic models are based on mathematically simple quadratic optimization criteria. And there is another thing to remember: A cybernetic system implements sparse coding among its variables. When looking at the resulting models, this means that relationships between variables are compressed and projected onto separate relations of the form $\bar{x}_i = q_i \bar{u}_i$. Complex systems are decomposable, only the action (input variable) and the reaction (corresponding latent variable) need to be studied at a time. Rather than having to manage the mess of all contributing variables at the same time, one can concentrate on a subset of localized variables.

Still, there are some points that deserve a closer look; what are the characteristics of this mathematics that is especially efficient in this discovery of new models?

Since the ancient Greeks, the methods of doing science have remained the same. The rules of the "game" were invented obeying some aesthetic and pragmatic objectives. For example:

> In Euclidean geometry, only a compass and a ruler were allowed in derivation of theorems, and there had to exist a finite sequence of operations for reaching the result. It turned out that a vast body of problems really could be formalized in this way – but, on the other hand, some problems turned out to be too difficult. Whereas an arbitrary angle could easily be divided in two equal parts, the seemingly analogous problem of *trisecting an angle* could never be accomplished. What is more, there were annoying inconsistencies: As the division in two and four equal parts was so simple, why the case there between them is so different?

For the Greek, mathematics was only a free men's pastime activity, not to be applied in real life. From the practical point of view, however, having a homogeneous conceptual toolbox that would work in all cases without abrupt collapses would be more useful. In the era of the computer, the rules of the game can be changed: *Infinite procedures* have become realistic. For example, employing this opportunity, trisecting the angle can be carried out by dividing the angle in two alternating halves an infinite number of times.

In the traditional way of thinking, the computational approach is not elegant — but it offers the homogeneity: Algorithms crunch the numbers, no matter what is the input data. The neocybernetic models simulating the operation of underlying agents, as implemented in the parallel fashion applying matrix methods, is robust but efficient — indeed, as Albert Einstein has said, "God does not care about our mathematical difficulties; he integrates empirically". Among mathematicians, there is resistance against computational approaches as the calculations cannot be carried out using pencil and paper, but a computer is needed. On the other hand, when seen in the neocybernetic perspective, the paradigm that trusts iterations towards convergence are very natural: Computation is also a process proceeding towards a dynamic balance.

Another change in mathematical thinking is most profound: In the neocybernetic spirit, relevance is more important than absolute truth or provability. When modeling large-scale complex systems, one can never assure that all assumptions are met that have to be fulfilled for some mathematical result to hold. This means that analyses can become either very misleading or completely meaningless because of sloppy simplifications. Even if the theories say that in certain circumstances the identification algorithms, say, converge to fixed values, this convergence can take infinite time — and sometimes it does, meaning that such methods are not practical. It is better to collect real information from the system and apply the models that are naturally dictated by the system; when data is gathered over longer time of system operation, and relevant structures — those that are visible in behaviors — in the data can be determined through statistical analysis. The same "sloppiness" also applies to today's methodologies: Different kinds of soft computing methods, etc., are today routinely used even though their operation cannot always be assured — it is enough that they usually work. Such practices can change the direction where the whole field is proceeding, as research is what the researchers do. The humans determine what is "hot" and what is not. When modeling ecosystems, for example, the dynamics of memetic system needs to be mastered just as well as the genetic one.

It should not be forgotten that a research community is a cybernetic system following its natural dynamics — and another complex cybernetic system with human minds as a medium is that of practicing engineers, those who finally either use the new approaches or not. One must not underestimate the role of intuition when estimating the dynamics in such system — completely new methodologies cannot easily penetrate. In this sense, neocybernetics nicely combines old and new ways of thinking: For example, in the field of industrial automation, the basic ideas are already quite familiar — control, information, ideas of local linearity, etc., are routinely employed.

There is also need for fresh ways of thinking what comes to the mathematical tools — but the need of fresh thinking goes deeper than that.

10.2.2 Questions of "why?"

Natural sciences like physics traditionally try to answer the "how?" questions: One derives formulas to explain behaviors and to estimate them. In biology and ecology, for example, the questions are of the form "what?", describing nature in terms of direct observations, and constructing taxonomies. In humanistic sciences the emphasis even seems to be on "who?", concentrating on singular cases rather than on statistical relevance, not to mention general rules. There is a hierarchy among the sciences what comes to their power in explaining phenomena — one could say that "how?" is an emergent-level problem setting as compared to "what?", and one cannot reduce such explanations to a set of lower-level ones in a one-to-one fashion. In the same manner, the next emergent level above "how?" is "why?".

If we trust Theodosius Dobzhansky, evolution must be seen as the basis of all biological and ecological phenomena. And the kernel of evolution boils down to this question.

It has been said that by definition science does not answer questions of that type — *teleology* and *finalism* are notorious words today. But what if this constraint could be relaxed; what if this starting point is just a problem of current ways of thinking, still reflecting the cybernetic battle between religious and non-religious explanations? After the studies in the Middle Ages, when all explanations *had* *to be* divine, they now *must not be.* After the other extreme we are now in the other, and balance is still being searched for. Indeed, these issues seem to be a *taboo* — so aggressive is the resistance against the creationistic movements, for example.

Still, it is clear that the old problem settings are becoming obsolete. Today, the data is so high-dimensional that there is an infinite number of ways how they can be explained — if only accepting the "how" explanations. There is never enough fresh data to cover the exponentially increasing space of the variables. Postmodern science is becoming a "fiddler's paradise" where the strangest formulas and theories are proposed. Observation data can be reduced into unequivocal formulas only if applying strong model structures and modeling principles — and the models based on the question "why?" give the tightest framework with the least number of free parameters. Instead of speaking of some *primus movens* or *elan vital* one can also more neutrally speak of *maximum entropy production* as studied in chapter 9; however, this is just a terminological trick, the fundamental underlying explanation for the behaviors remaining equally mysterious.

In technical terms, nothing very peculiar takes place: Finalistic criteria are often reflected as optimization problems. The motivation for this is that operation in a cybernetic system is based on the competition among the low-level agents — as seen from outside, in the slower scale, the "losers" cannot any more be seen, only "winners" being visible. The Fermat's principle (light chooses the fastest path) can be explained in terms of photon populations; similarly, the idea of the "selfish gene" [20], for example, can be reduced to analysis of populations. The way of looking at the emergent illusions just makes the system look clever and behaviors in them look pre-planned; finalistic arguments make it possible to express the emergent patterns in a compact form.

When doing cross-disciplinary studies, one must not underestimate the role of intuitions and connotations there are. After all, however, it does not matter whether one speaks of emergy pressure or tension — or of "life force" or "will", in the sense of Arthur Schopenhauer.

There also exist other kinds of emergent-level criteria for constructing models. Remembering that our own mental machinery has been polished by millions years of evolution, one can perhaps also propose "anthropomimetic modeling". What kind of dependency structures does the mental machinery utilize in its subconscious modeling process? Using everyday terminology, it seems that one of such principles is *beauty,* or, more specifically, *symmetry.* Symmetries make it possible to capture dependencies in patterns, thus efficiently compressing information, as studied in chapter 9. Harmonic patterns please the eye because they match our innate models — or they help to enhance the models. Extending this to other systems is not as esoteric as it sounds — why should the human be the only example of natural systems where such sophisticated models of fractal symmetries are introduced; perhaps nature is beautiful of necessity, to make the quantum-level systems cybernetically modellable and thus controllable? Perhaps nature is understandable not only at the conscious level, but also on the very deep sub-conscious level — but, still, this beauty is there not to please us; it may be that an ugly universe could not be balanced in the first place! Anyway, symmetries are more or less explicitly used as the starting point of modeling in quantum physics today — mainly for pragmatic reasons, as this is the only way to decrease number of parameters there, but perhaps these symmetries really reflect reality. It is not only the best of possible worlds, in the sense of Leibnitz (and also in the neocybernetic sense), but it is also the most beautiful one — Miss Universe, as judged by the jury of systems (including us).

Perhaps there are still more general ideas available — guidelines that could be exploited to model and control the higher-level memetic system of *science making* itself. When trying to understand the world, and when modeling it scientifically, one is facing data coming from complex systems. But if the system being modeled is cybernetic, one already knows that there exists a model — the model that is implicitly being used by nature for keeping that system hold together. Here it is reasonable to present the *Gaia hypothesis* by Joe Lovelock:

> The Goddess of Earth (Gaia) has purposefully designed the geological, climatological, etc., processes to support life.

This strange principle makes it possible to draw strong conclusion concerning life on earth; there also exist closer analyses (for example, see [45], [33]). Similarly, one can propose such emergent-level laws — let us define the "Pallas Athene Hypothesis", or, actually, the "Antero Vipunen Hypothesis", as follows:

> The God(dess) of Science has purposefully designed the biological, ecological, etc., processes to support science.

The triumph cannot end now — the Goddess of Science protects us so that nature will continue to be graspable to humans, and science will continue to prosper. It seems that system complexity and analyzability go hand in hand: If nature has been able to construct sophisticated model structures for the cybernetic systems, why not us? The claim here also is that *cybernetic systems can always be modeled,* human's task is to detect these models. When searching for the models, there are always many ways to proceed; but the above hypothesis gives strict guidelines about where to go next; indeed, this whole text is based on such assumptions — or the world is seen through "neocybernetic eye-glasses". One of such fundamental assumptions was that of (piecewise) linearity: Otherwise, reductionism does not work. However, even though the model structures are linear, it is not so that the combined system would be simply decomposable: As new boundaries against outer environment are exploited, new "forces" are detected, and the "steel plate" is no more the same. But when one starts from simple assumptions, the result remains simple — the linearity assumption is like the *parallel axiom*: if this assumption is adopted, a consistent, nontrivial framework can still be derived, revealing a very different view of the world.

Actually, the key question in neocybernetics is not "why?", but "why not?".

10.2.3 Mental traps

Many beliefs from the past seem ridiculous to us — how about *our* beliefs as seen hundreds of years in the future? Even though we know so much more than the medieval people, it is difficult to imagine what we cannot yet imagine. And, indeed, because of the new measurement devices and research efforts, the number of "non-balanced" observations and theories is now immense. There are many fallacies and logical inconsistencies in today's top science — many of these are related to the astonishingly clever orchestration and implicit control of complex processes. Categorically avoiding the "why?" questions results in unsatisfactory models: Today's explanations for gene transcription and translation, for example, make the role of message-RNA sound, really, like an agent story, this agency representing some central intelligence.

Today's challenges involve different kinds of complex networks, and today's conceptual tools do not help in understanding them. Indeed, when seen in the appropriate perspective, most challenges today can be characterized in terms of networked distributed agents: For example, the advertised possibilities of nanotechnology can only become true if some mechanism of self-organization and self-regulation among the nano-things can be implemented.

The agents doing science are humans, and it is the patterns of "common sense" thinking that have to be overcome to reach new ways of seeing the world. A person that has adopted Western thinking is in the prison of centralized thinking. To understand complex cybernetic systems, old thinking patterns have to be recognized. It is like the Zen masters say, "if you are thinking of Buddha, you must kill it" — or if you notice that you are thinking you must stop it! It is the religious ideas that are among the most fundamental patterns of thought, one of them in the Western culture being the monotheist dogma of a personalized creator. After detecting such thinking patterns, escaping them is still not easy. Jean-Paul Sartre has said that *even the most radical irreligiousness is Christian Atheism* — one explicitly (aggressively) tries to eliminate all divine-looking explanations, but in vain.

Getting distributed by definition means loosening the controls. Indeed, traditional centralized control is a prototype of Western ways of structuring and mastering the world. But even though the Eastern wisdom better captures the essence of cybernetic phenomena, engineering-like approaches are necessary to "bootstrap" ones understanding, and powerful conceptual tools are needed to construct new kinds of emergent structures. The hermeneutic mind alone is not enough, one needs a link to the outer world. Some kind of synthesis of Eastern and Western ways of thinking is needed; and neocybernetics seems to offer a framework where it is possible to reductionistically study holistic distributed systems.

Can "cybernetics" then offer some alternative content for one's personal world? Fortunately, creating maxims out of nothing is not needed — it has already been done. Eastern holistic thinking offers a model of how to create the new world view. For example, the underlying vitality principle beyond the Chinese philosophy and medicine is based on the ideas of balancing and ordering (see Fig. 10.3); in Indian philosophy many principles (stationarity, desire and consequent suffering, etc.) reflect the cybernetic ideas; and the Japanese pantheistic belief on the millions of Gods, each managing its own subsystem, is also appealing. According to Eastern philosophers, the reason for suffering is missing knowledge and understanding.

Figure 10.3: Chinese symbol for the mystical *ordering principle,* also denoting *air* or *vapor*

10.3 Rehabilitation of engineering

In the beginning, in chapter 1, it was wondered whether control engineering can have anything to do with biology. As a conclusion, it can be claimed that there is contribution in both directions.

Since 1960's, after the great discoveries of modern control theory, there have been no real breakthroughs in the branch of control engineering. It seems that this stagnation does not need to last long: There is a Golden Age of control engineering ahead. Control theory and tools can be applied not only in technical applications, but also in understanding really complex system — biological, social, economical, etc. There do not necessarily exist explicit controls in such systems, but understanding the natural dynamics in such systems is still based on control intuitions.

It is traditionally thought that philosophy is the basis of all science: Logic is part of philosophy determining the rules of sound thinking. Mathematics if "applied logic", implementing the logical structures and manipulating them according to the logical rules. Natural sciences, on the other hand, can be seen as "applied mathematics", where the ready-to-use mathematical formulas are exploited to construct models. Finally, the engineering disciplines are "applied science". Engineering is inferior to the more fundamental ways of structuring the world.

This is a formal view of how deductive science is done, and how new truths are derived. However, also these viewpoints need to be reconsidered: If the presented neocybernetic modeling can cast some light onto the mysteries of what is the essence of complex systems, the deepest of the philosophical branches, metaphysics, is addressed. It is mathematics that offers the *syntax* for discussing the issues of what is there beyond the observed reality, and it is control engineering that offers the *semantics* into such discussions. It can be claimed that control knowledge is necessary for understanding complex systems, natural or artificial (see Fig. 10.4).

Martin Heidegger once said that as classic philosophy fades away, cybernetics becomes a philosophy for the twentieth century. This will be even more true in the 21'st century: Philosophical considerations can be interpreted from the viewpoint of information science, they can be given fresh contents, and, what is more, there are new conceptual tools available — the language of mathematics.

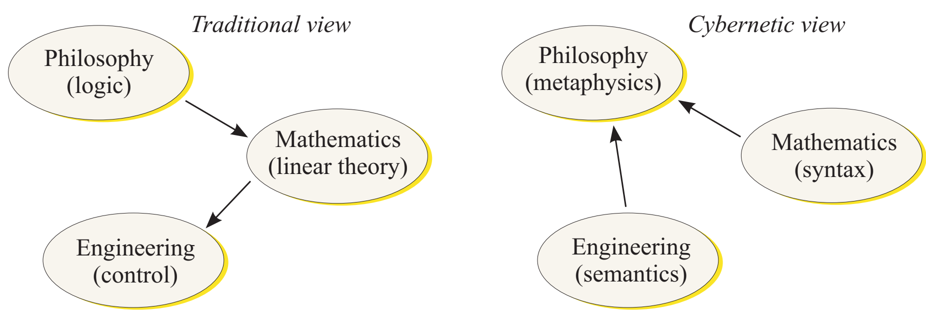


Figure 10.4: New view of control engineering studies as delivering substance to philosophies

... Beyond the Level 11

Conclusion[1]: From Science back to *Natural Philosophy*

When studying memetic systems, one can see that there are emergent hierarchies. For example, artificial intelligence research has strange appeal: It seems to be always *ahead of paradigms.* If a concrete formulation is found for some AI problem, it can be implemented by hard work; it is no more *interesting* — and it is no more AI. When a study already has form and fixed paradigm, standard methods and problems, it becomes a memetic system of its own. Similarly, there is a category above all sciences, defying scientific study — *we just know it exists.*

11.1 Standard science — business as usual

Good science — this is one of the main goals in today's universities. What is the definition of "good science", then? Indeed, today science is measured using very concrete productivity criteria. Researchers and project proposals are evaluated using panels, peer reviews, and different kinds of publication indices. This information is utilized to redirect financing, for "focusing on the strengths". Who could oppose efficiency?

There also exist different kinds of development efforts to enhance efficiency in universities. There are questionnaires mapping the working practices, and new planners and analysts are hired to implement the "missions" and "visions". New practices are introduced, including "near-bosses", "developmental discussions", etc, making matrix organizations hierarchical again. In short, information acquisition processes are intensified, and system controls are adapted accordingly.

As the system becomes better measured and more efficiently controlled, the system becomes *cybernetized,* as studied above. This means that the number of degrees of freedom decrease, the system is better predictable and deviations from the nominal are minimized.

But what is that "nominal" in science? In the Kuhnian terminology, it is as-

[1]The conclusions here obviously do not reflect the opinions of the University, or those of the Department, or those of the Laboratory

sumedly "standard science". One should be searching for something new that nobody knew before — but for such unknown thing there cannot exist measurements and no controls. Doing science does not match the efficiency pursuit. To survive, a researcher has to compromise: One has to trivialize the problems, searching for "easy wins", making his/her achievements better quantifiable and predictable. Clever people adapt, optimizing locally, producing the stuff that is being required. Diversity is effectively eliminated from the system.

According to the neocybernetic discussions, the system becoming cybernetized ends in *stagnation,* free flow of thoughts changing to pre-programmed bureaucracy. But what is even more alarming is that there is *loss of vitality.* Enthusiasm is necessary in science[2]. By making the scientific practice non-appealing, the brightest minds select other careers — they usually have the choice. Cynicism and pessimism are very acute threats for loss of interest that gnaw the memetic system from inside. The potential for breakthroughs is minimized, still worsening the vicious circle of systemic degeneration.

Where is the opposition, counterarguments that would introduce some noise and excitation in the system, preventing it from ending in stagnation? It has to be recognized that there are powerful pressures keeping the *status quo.* The arguments often become personified, and nobody wants to disagree with the celebrated top scientists, those who have the aura of heavenly wisdom — and who would not like the world to change. The general atmosphere is discouraging, as it is thought that the "backward-looking traditionalist" just "do not understand". There is too much to lose for a person trying to make a career. It is the same problem with "scientific spirit" as it is with "free will" — people do not want to take the responsibility, after all. Is there then any hope?

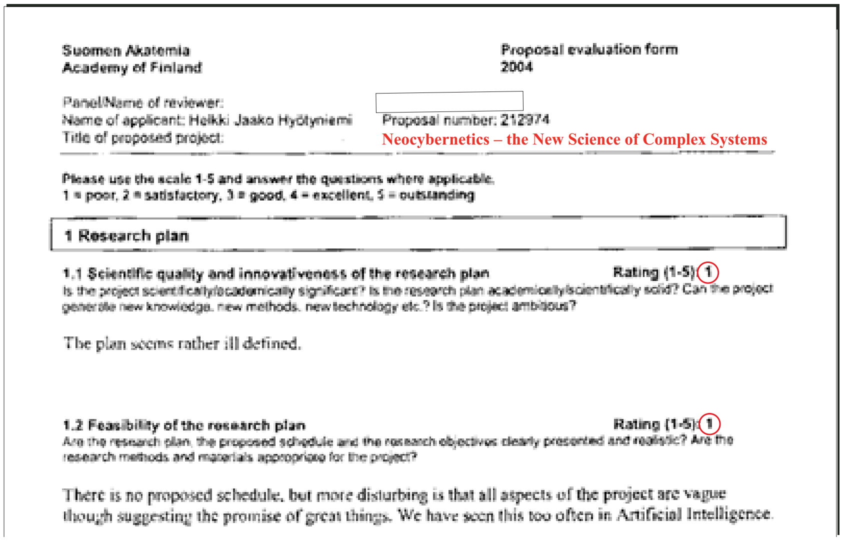
11.2 "Project 42"

In some form science will always survive, even though today's ways of doing it can collapse or degenerate. One needs to look at science in a wider perspective — or, more generally, one should speak of *natural philosophy.* Natural philosophy is the higher-level category hosting different kinds of incarnations of science. It seems that cybernetization in sciences cannot be avoided, but after catastrophes, new ways of doing science displace old ones.

After all, Isaac Newton was not a scientist: According to his own words, he was a natural philosopher. Natural philosophy leaves mature, gradually paralyzing sciences along the path of its ever-proceeding Geist.

But the above criticism about today's science only applied to the framework, not the actual substance — is there need for the contents of scientific paradigms to change? It seems that regeneration truly is necessary. Richard Feynman has claimed that *one should not even try to understand quantum phenomena.* The best available theory today, or *quantum electrodynamics,* gives good predictions, but offers no intuition into the world of elementary particles. Why should one be satisfied with such unsatisfactory models? The purpose of science is not only to carry out calculations, but also to reach understanding.

[2] As one astronomy graduate preparing her Doctoral Thesis lately confessed: "A trained monkey could also type these figures in the computer"


Suomen Akatemia
Academy of Finland

Proposal evaluation form
2004

Panel/Name of reviewer:
Name of applicant: Heikki Jaako Hyötyniemi
Proposal number: 212974
Title of proposed project: Neocybernetics – the New Science of Complex Systems

Please use the scale 1-5 and answer the questions where applicable.
1 = poor, 2 = satisfactory, 3 = good, 4 = excellent, 5 = outstanding

1 Research plan

1.1 Scientific quality and innovativeness of the research plan — Rating (1-5): 1
Is the project scientifically/academically significant? Is the research plan academically/scientifically solid? Can the project generate new knowledge, new methods, new technology etc.? Is the project ambitious?

The plan seems rather ill defined.

1.2 Feasibility of the research plan — Rating (1-5): 1
Are the research plan, the proposed schedule and the research objectives clearly presented and realistic? Are the research methods and materials appropriate for the project?

There is no proposed schedule, but more disturbing is that all aspects of the project are vague though suggesting the promise of great things. We have seen this too often in Artificial Intelligence.

Figure 11.1: Official evaluation of the "neocybernetics" ideas back in 2005 (excerpt). The main purpose of the proposed project would have been to complete a monograph on "Neocybernetics in Biological Systems"

In the "Project 42" the goal is to find models for life and universe — for all complex systems[3]. And these models should be simple: The sincere belief is that nature can be understood by a human. As an application, the goal is to detect processes of cybernetization — and fight against it in those domains where it is not suited.

It may be that this research is not good science. Indeed, this has been indicated indisputably by the highest authority, the Academy of Finland (see Fig. 11.1). But perhaps this is still good natural philosophy? As Edward Goldsmith puts it when discussing his thoughts concerning Deep Ecology [33]:

> ... Our mainstream biologist, ecologists and anthropologists — will certainly reject them. I hope they do. If they do not, then I know that the laws must be seriously wanting, for I regard today's mainstream natural sciences (biology, ecology and anthropology) as being very seriously misguided ...

11.3 Neocybernetics — an experiment design

Experiment designs in complex systems are difficult to carry out, and proving hypotheses in memetic theories is practically impossible. The goal here is to test whether the cybernetic ideas hold, and how the attractors in the memetic sphere emerge and how they find their balance.

[3]According to "The Hitchhiker's Guide to the Galaxy", the trilogy in five parts by Douglas Adams, the Definite Answer to the Ultimate Question of Life, the Universe and Everything is **42**. Only the question is inaccurate

There exist no proofs for theories in complex enough domains. Verification of claims, on the other hand, is implemented by checking whether they can pass the credibility and relevance test. Ways of doing science change: Cybernetic proof techniques are not based on truth but vitality, the capacity of the ideas to compete and stay alive. If the theory passes such test, it has to capture some essence of the real system *as we see it.* In the spirit of cybernetics, the proof and the theory itself are intertwined and also relative to the context.

This text is a *cybernetic proof of itself,* or it remains a "proof" of the contrary, and the readers of this text are the agents implementing the *emergent proof.*

Dear reader: If this text has had the momentum for some reason to reach your knowledge just due to its own virtues, bypassing all scientific authorities, being (seemingly) incompatible with today's active scientific memes — then *it must be relevant* (not claiming anything about its final *truth*).

On the other hand, if you *do not* ever bump into this text, you should be happy in your ignorance: It was then probably not worth knowing in the first place, it would have been only waste of time.

Seamless information transfer and its more homogeneous penetration is a prerequisite for science. This text is available in Internet — as it can be freely downloaded, it hopefully finds its "memetic balance" in the ideasphere all by itself. Scientific theories must always be based on cybernetic tensions among arguments and counterarguments — I would be very happy if somebody would propose what is the contents of level 11 and onwards in the ladder of deeper cybernetic understanding. As Heraclitus and Hegel once observed, the key point is not *being* but always *becoming* — perhaps the presented ideas help to smoothen the transition to something qualitatively new.

If you have read this text and found it *interesting* and *understandable* (which are, after all, the most relevant criteria for memes to survive in human minds), I would be happy if you would send a note to `heikki.hyotyniemi@tkk.fi`. Thank you for your interest!

Suuni jo sulkea pitäisi
kiinni kieleni sitoa
laata virren laulannasta
heretä heläjännästä

...

Vaan kuitenki, kaikitenki
laun hiihin laulajoille
laun hiihin, latvan taitoin
oksat karsin, tien osoitin
Siitäpä nyt tie menevi
ura uusi urkenevi
laajemmille laulajoille
runsahammille runoille
nuorisossa nousevassa
kansassa kasuavassa.

– *Kalevala*

Bibliography

[1] J.R. Anderson: *The Architecture of Cognition.* Harvard University Press, Cambridge, Massachusetts, 1983.

[2] K.J. Åström and P. Eykhoff: System identification — A survey. *Automatica,* Vol. 7, pp. 123–162, 1971.

[3] K.J. Åström and B. Wittenmark: *Adaptive Control.* Addison-Wesley, Boston, MA, 1995 (second edition).

[4] K.J. Åström and B. Wittenmark: *Computer-Controlled Systems.* Prentice Hall, NJ, 1997.

[5] A.-L. Barabasi: *Linked: The New Science of Networks.* Perseus Publishing, 2002.

[6] A. Basilevsky: *Statistical Factor Analysis and Related Methods.* John Wiley & Sons, New York, 1994.

[7] G. Bateson: *Steps to an Ecology of Mind.* Paladin Books, 1973.

[8] J.A. Berson: *Chemical Creativity: Ideas from the Work of Woodward, Hckel, Meerwein, and Others.* Wiley, 1999.

[9] L. von Bertalanffy: *General System Theory — Foundations, Development, Applications.* George Braziller, New York, NY, 1969 (revised edition).

[10] Brehm, J.J. and Mullin, W.J.: *Introduction to the Structure of Matter.* John Wiley & Sons, 1989.

[11] R.A. Brooks: *Cambrian Intelligence.* MIT Press, 1999.

[12] A. Bunde and S. Havlin (Eds.): *Fractals in Science.* Springer Verlag, 1994.

[13] F. Capra: *The Web of Life.* Anchor Books, New York, 1996.

[14] J.M. Carlson and J. Doyle: Highly Optimized Tolerance: Robustness and Design in Complex Systems. *Physics Review Letters,* Vol. 84, No. 11, pp. 2529–2532, 2000.

[15] H. C. Causton, *et al.*: Remodeling of Yeast Genome Expression in Response to Environmental Changes. *Molecular Biology of the Cell,* Vol. 12, pp. 323–337, February 2001.

[16] F.E. Cellier: *Continuous System Modeling.* Springer-Verlag, New York, 1991.

[17] W.G. Chase and H.A. Simon, "The minds eye in chess". In W. Chase (ed.), *Visual information processing.* Academic Press, New York, 1973.

[18] Chomsky, N.: *Syntactic Structures.* Mouton, The Hague, 1957 (Reprint Berlin and New York, 1985).

[19] R.D. Cook, D.S. Malkus, M.E. Plesha, and R.J. Witt: *Concepts and Applications of Finite Element Analysis.* Wiley & sons, 2001 (4th edition).

[20] R. Dawkins: *The Selfish Gene.* Oxford University Press, 1976.

[21] R. Dawkins: *The Blind Watchmaker.* Penguin Books, London, 1991.

[22] J. Diamond: *Guns, Germs, and Steel: The Fates of Human Societies.* W.W. Norton & Co, New York, 1997.

[23] F.D. Dyson: *Origins of Life.* Cambridge University Press, New York, 1985.

[24] G.C. Dean: An Introduction to Kalman filters. *Measurement and Control,* Vol. 19, pp. 69–73, 1986.

[25] H. Lindstone and M. Turoff (eds.): *The Delphi Methods.* Addison-Wesley, Boston, MA, 1975.

[26] K.I. Diamantaras and S.Y. Kung: *Principal Component Neural Networks: Theory and Applications.* Wiley, New York, 1996.

[27] J.S. Edwards, R. Ramakrishna, C.H. Schilling, and B.O. Palsson: Metabolic Flux Balance Analysis. In S.Y. Lee and E.T. Papoutsakis (eds.) *Metabolic Engineering.* Marcel Decker, pp. 13–57, 1999.

[28] C. Emmeche: *The Garden in the Machine. The Emerging Science of Artificial Life.* Princeton University Press, 1994.

[29] P. Földiák: Sparse coding in the primate cortex. In M.A. Arbib (ed.): *The Handbook of Brain Theory and Neural Networks.* MIT Press, Cambridge, MA, 2002 (second edition).

[30] S. Franchi and G. Gzeldere (eds.): *Constructions of the Mind: Artificial Intelligence and the Humanities.* A special issue devoted to the *exploration of convergences and dissonances between Artificial Intelligence and the Humanities* of the Stanford Humanities Review, Vol. 4, Issue 2.

[31] P. Gärdenfors: *Conceptual Spaces.* MIT Press, 2000.

[32] A. P. Gasch, *et al.*: Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes, *Molecular Biology of the Cell,* Vol. 11, pp. 4241–4257, December 2000.

[33] E. Goldsmith: *The Way: An Ecological World View.* University of Georgia Press, Athens, Georgia, 1998.

[34] O. Haavisto and H. Hyötyniemi: Data-based modeling and control of a biped robot. *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'05,* Helsinki, Finland, pp. 427–432, June 27–30, 2005.

[35] O. Haavisto, H. Hyötyniemi, and C. Roos: State space modeling of yeast gene expression dynamics. *To be submitted.*

[36] S. Haykin: *Neural Networks — A Comprehensive Foundation.* Prentice–Hall, Upper Saddle River, NJ, 1999.

[37] D.O. Hebb: *The Organization of Behavior: A Neuropsychological Theory.* John Wiley & Sons, New York, 1949.

[38] J.H. Holland: *Hidden Order: How Adaptation Builds Complexity.* Addison-Wesley, Boston, MA, 1996.

[39] N.S. Holter, A. Maritan, M. Cieplak, N.V. Fedoroff, and J.R. Banavar: Dynamic modeling of gene expression data. *PNAS,* Vol. 98, No. 4, pp. 1693–1698, February 2001.

[40] J. Horgan: *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age.* Helix Books, New York, NY, 1997.

[41] A. Hyvärinen, J. Karhunen, and E. Oja: *Independent Component Analysis.* John Wiley & Sons, New York, 2001.

[42] H. Hyötyniemi: *Multivariate Regression — Techniques and Tools.* Helsinki University of Technology, Control Engineering Laboratory, Report 125, 2001.

[43] S. Johnson: *Emergence: The Connected Lives of Ants, Brains, Cities, and Software.* Touchstone, 2002.

[44] S.A. Kauffman: *At Home at the Universe.* Oxford University Press, New York, 1995.

[45] A. Kleidon, R.D. Lorenz (eds.): *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond.* Springer–Verlag, Berlin, 2004.

[46] T. Kohonen, T. *Self-Organizing Maps.* SpringerVerlag, Berlin, 2001.

[47] B. Kosko: *Fuzzy Thinking: The New Science of Fuzzy Logic.* Hyperion/Disney Books, 1993.

[48] C.J. Krees: *Ecology. The Experimental Analysis of Distribution and Abundance.* Benjamin Cummings, San Francisco, 2001.

[49] T. Kuhn: *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago, 1962.

[50] J. Laaksonen: *Subspace Classifiers in Recognition of Handwritten Digits.* Acta Polytechnica Mathematica, Mathematics, Computing and Management in Engineering series, No. 84, Espoo, 1997.

[51] V. Lesser, C.L. Ortiz Jr., and M. Tambe (eds.): *Distributed Sensor Networks — A Multiagent Perspective.* Kluwer Academic Publishers, Boston, MA, 2003.

[52] B.D. Malamud, G. Morein, and D.L. Turcotte: *Forest Fires: An Example of Self-Organized Critical Behavior. Science,* Vol. 281, Issue 5384, pp. 1840–1842, September 18, 1998.

[53] H. Maturana and F. Varela: *Autopoiesis and Cognition.* D. Reidel, Dordrecht, Holland, 1980.

[54] H.H. McAdams and L. Shapiro: Circuit simulation of genetic networks. *Science,* Vol. 269, pp. 651–656, 1995.

[55] M. Mitzenmacher: A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics,* Vol. 1, No. 2, pp. 226–251.

[56] Morrison, R.T. and Boyd, R.N.: *Organic Chemistry.* Allen and Bacon Inc., 1987 (5th edition).

[57] J.D. Murray: *Mathematical Biology. Part 2: Spatial Models and Biomedical Applications.* Springer, New York, third edition 2002.

[58] H.T. Odum: *Environmental Accounting, Emergy and Decision Making.* John Wiley, New York, 1996.

[59] B.A. Olshausen and D.J. Field: Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research,* Vol. 37, pp. 3311–3325, 1997.

[60] P. van Overschee and B. de Moor: *Subspace Identification of Linear Systems.* Kluwer Academic Publishers, 1996.

[61] M. Parkin: *Microeconomics.* Addison Wesley Longman, 2004 (7th edition).

[62] J. Pearl: *Causality: Models, Reasoning, and Inference.* Cambridge University Press, Cambridge, MA, 2000.

[63] R. Penrose: *The Emperor's New Mind.* Oxford University Press, 1990.

[64] I. Prigogine: *End of Certainty.* The Free Press, 1997.

[65] R. Rohde and R. Muller: Cycles in fossil diversity. *Nature,* **434**, 2005.

[66] E. Rosch: Principles of Categorization. In E. Rosch and B.B. Lloyd (eds.): *Cognition and Categorization.* Erlbaum, Hillsdale, NJ, 1978.

[67] S. Russell and P. Norvig: *Artificial Intelligence: A Modern Approach.* Prentice-Hall, 1995.

[68] T.L. Saaty: *The Analytic Hierarchy Process.* McGraw-Hill, New York, 1980.

[69] E. Schrödinger: *What Is Life?* Macmillan, 1947.

[70] J. Searle: Minds, Brains, and Programs. *Behavioral and Brain Sciences,* Vol. 3, pp. 417–424, 1980.

[71] Senge, P.M.: *The Fifth Discipline: The Art and Practice of the Learning Organization.* Doubleday Currency, New York, 1990.

[72] H.A. Simon: *Sciences of the Artificial.* MIT Press, Cambridge, MA, 1996 (third edition).

[73] M. Sipser: *Introduction to the Theory of Computation.* Course Technology, 2005 (second edition).

[74] B. van Steensel: Mapping of genetic and epigenetic regulatory networks using microarrays. *Nature Genetics,* Vol. 37, pp. S18–S24, 2005.

[75] D.W. Stephens and J.R. Krebs: *Foraging theory.* Princeton University Press, 1986.

[76] J.D. Sterman: *Business Dynamics — Systems Thinking and Modeling for a Complex World.* Irwin McGraw-Hill, Boston, MA, 2003.

[77] R. Tarnas: *The Passion of the Western Mind — Understanding the Ideas That Have Shaped Our World View.* Ballantine Books, New York, 1991.

[78] A.H.Y. Tong, *et al.*: Global Mapping of the Yeast Genetic Interaction Network. *Science,* Vol 303, Issue 5659, pp. 808-813, February 6, 2004.

[79] P. Turchin: *Complex Population Dynamics: A Theoretical/Empirical Synthesis.* Princeton University Press, 2003.

[80] A. Turing: Computing Machinery and Intelligence. *Mind,* Vol. LIX, pp. 433–460, 1950.

[81] A. Turing: The Chemical Basis of Morphogenesis. *Philosophical Transactions of The Royal Society: Biological Sciences,* Vol. 237, pp. 37–72, 1952.

[82] J. Vohradský: Neural network model of gene expression. *FASEB Journal,* Vol. 15, March 2001.

[83] F. de Waal: *Our Inner Ape.* Riverhead Books, New York, 2005.

[84] M.S. Waterman: *Introduction to Computational Biology — Maps, Sequences and Genomes.* Chapman & Hall, London, 1995.

[85] E. Weitzke and P.J. Ortoleva: Simulating cellular dynamics through a coupled transcription, translation, metabolic model. *Computational Biology and Chemistry,* Vol. 27, pp. 469–480, 2003.

[86] N. Wiener: *Cybernetics: Or Control and Communication in the Animal and the Machine.* Wiley, New York, NY, 1948.

[87] E. Wigner: The Unreasonable Effectiveness of Mathematics in the Natural Sciences. *Communications in Pure and Applied Mathematics,* Vol. 13, No. 1, 1960.

[88] C. Wills and J. Bada: *The Spark of Life: Darwin and the Primeval Soup.* Perseus Publishing, Cambridge, MA, 2001.

[89] E.O. Wilson: *The Diversity of Life.* Harvard University Press, 1992.

[90] E.O. Wilson: *Consilience: The Unity of Knowledge.* Abacus, 1999.

[91] S. Wolfram: *A New Kind of Science.* Wolfram Media, Champaign, IL, 2002.

[92] *Additional material on neocybernetics is available in public domain at the Internet address* `http://www.control.hut.fi/cybernetics`.

Because of the cross-disciplinary nature of cybernetic studies, the above list of references is not exhaustive. Omissions of some relevant material is not intentional: As far as the author can tell, the ideas and models of neocybernetics are anyway original.