ENFORMATION THEORY

"Physical Vitalism"

Heikki Hyötyniemi

Presentation in Jyväskylä, April 26, 2013
**Presentation in Helsinki, May 6, 2013**

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

Heikki Hyötyniemi

- Professor of automation technology at Helsinki University of Technology between 2001–2009
- Background in artificial intelligence and neural networks (specially self-organizing maps)

  - Later … from science nearer to more general *natural philosophy*?

ENFORMAATIOTEORIA
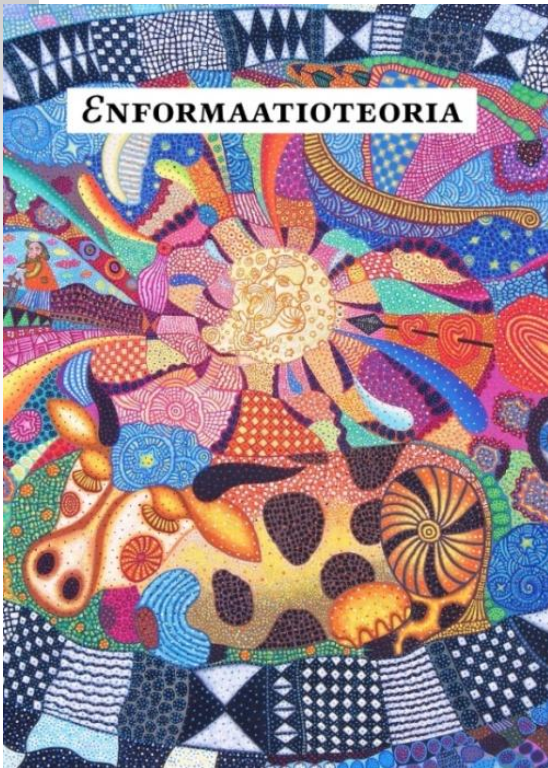
heikki.hyotyniemi@aalto.fi

Collaboration with *Petri Lievonen*

- Man behind the scenes – ideas, corrections, simulations, courses, web pages, links…
- Taking a more visible role in the "enformation theory future"?

petri.lievonen@hiit.fi

Binary data

Linear neurons

Activation function tanh(x)

Activation function cut(x)

Activation function |x|
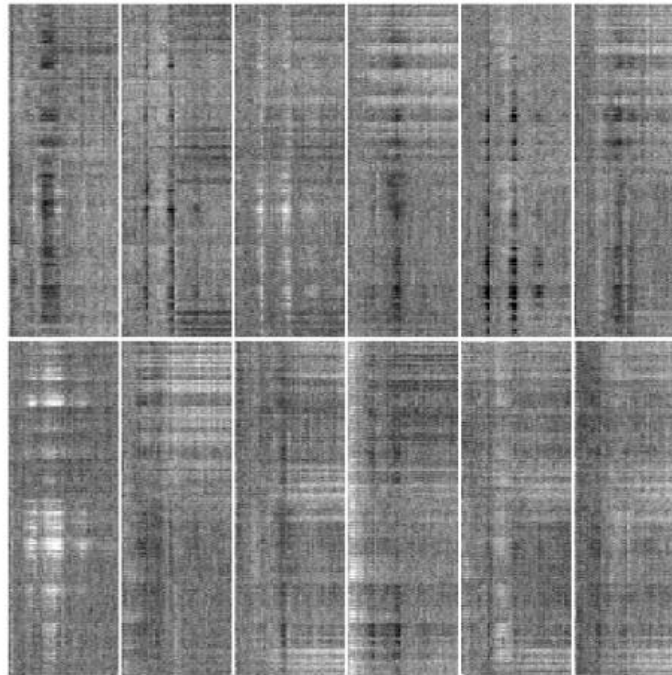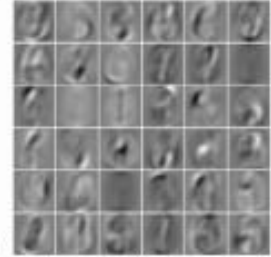
Activation function log(cosh(x))

Figure A.6: Profiles of 12 emergent components of MEG data, now augmented with class information in extra dimensions. This external information rotates and mixes the profiles to facilitate classification. With these linear components one can predict with 60% accurary which of the five classes the subject is watching.
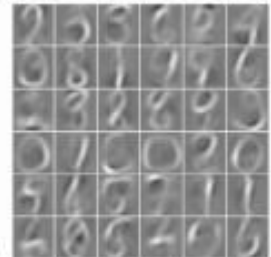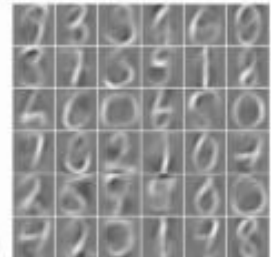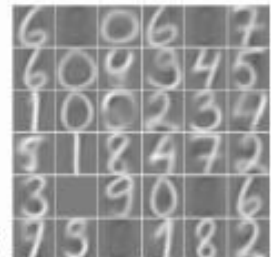
Beyond first principles

- Pure *empirism* is becoming challenged: There is too much data, too many interpretations

- Empirism should be combined with *rationalism* to reduce the degrees of freedom in data

- One would like to have some general *a priori* model structure for measurement data that exists

- … But assuming some structure for the world is *philosophy* – even *metaphysics*!

- Today's *analytical philosophy* sees "world as a set of facts" (Wittgenstein's early work)

- A fresh and intuitive approach is to try and *escape the static world view* (W's later work)
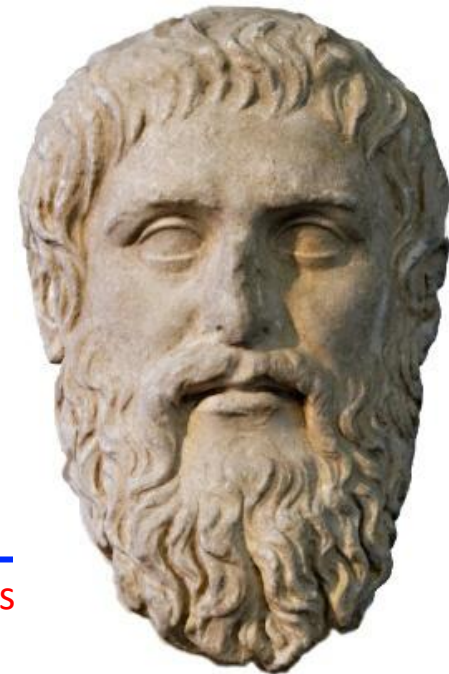
Heraclitus

Information from chaos?

- *Process philosophy* emphasizes the dynamic essence of the world – *becoming* is more important than *being*

- Assumption now: Everything that is relevant is *dynamics*, in its simplest form only some kind of *vibration*

- Later … it turns out that correlating vibrations constitute *flows* and *dynamic attractors* therein

- So, *process* and *dynamics* are more relevant than *structures* or *mechanisms*

- To simplify things to the extreme, assume that there is *no structure whatsoever* to begin with

- Start from the "birth", from complete chaos, and study what kind of *information* (data distributions) there is available

"VITALISM" without FINALISM

No structure – *normal* (*Gaussian*) *distribution*

Probability $N\left\{ \mathrm{E}\{z_i\},\ \mathrm{E}\{z_i^2\} - \mathrm{E}\{z_i\}^2 \right\}$

"Static world"

"Forgotten"?

*This is **not** only measurement noise*

*This is the **vibration***

$\sqrt{\ }$

spread

expectation

z_i
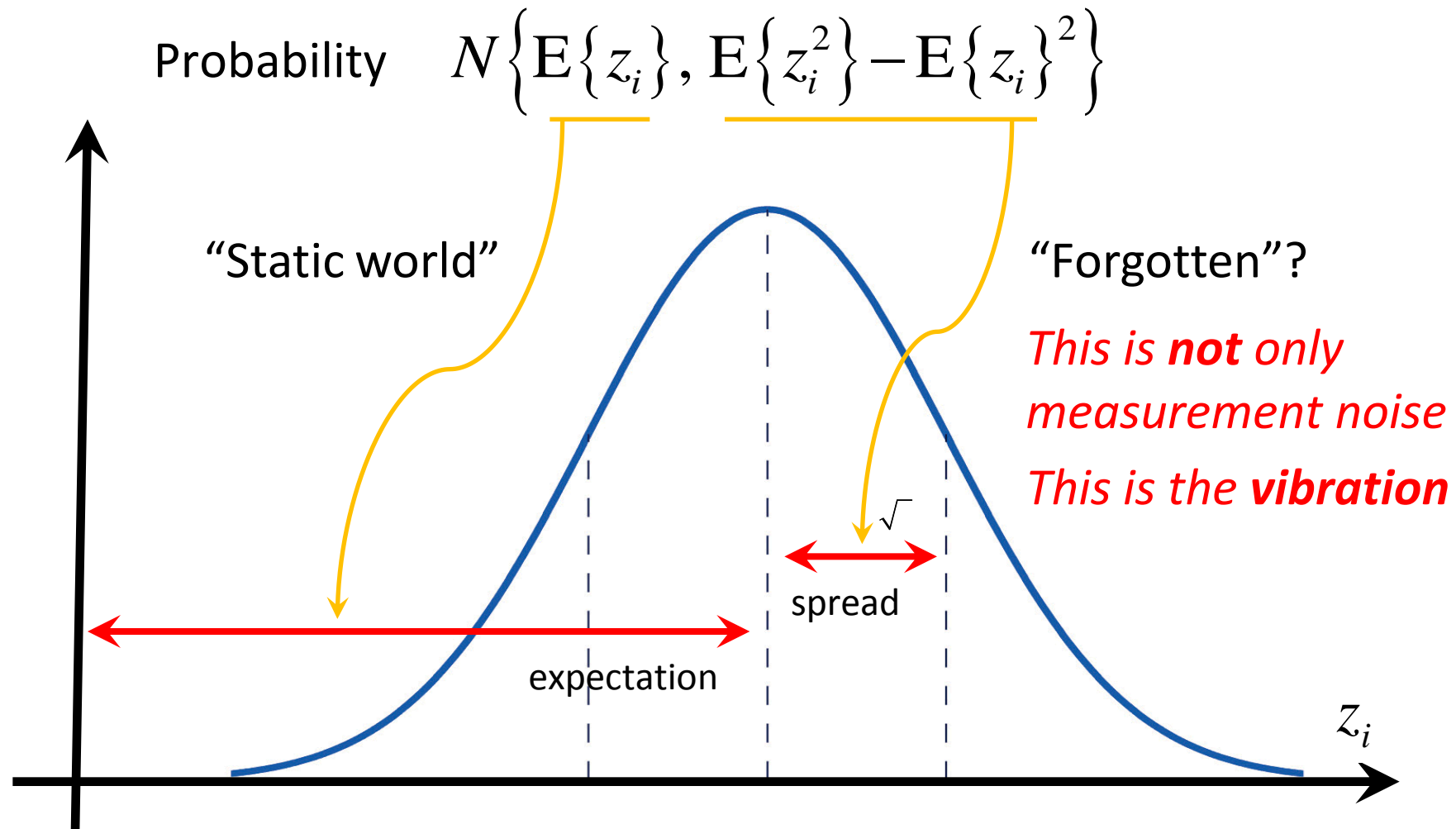
Becomes interesting in higher dimensions, in the case of *multivariate distributions*

Characterizing distributions – *interpretations*

- *"Matter"* – observed reality; averaging, *coarsening*

$$\mathrm{E}\{z_i\}$$

- $\boxed{Enformation}$ – variation, nature of change, *dynamics*

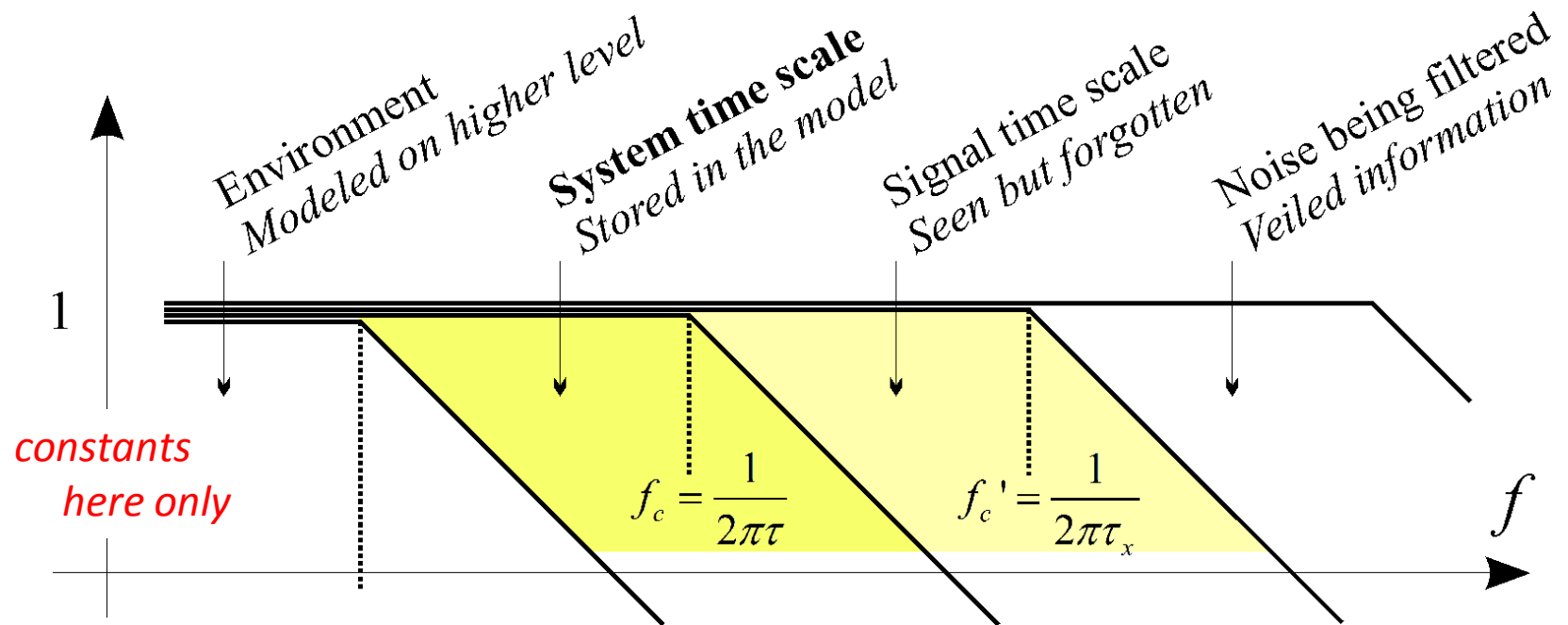$$\mathrm{E}\{z_i^2\} - \mathrm{E}\{z_i\}^2 \quad \text{variance} \quad \textit{correlations – covariances}$$

- "If everything is *springs* and the state of a spring (deviation from balance / amplitude) is z_i, then its *energy* is proportional to z_i^2"
- "Energetic information" – capacity to change the world = basis for *natural semantics*
- It turns out that enformation can be interpreted as *vital force* ?

"HOW" ⟶ **"WHY"**

Practical enformation

Environment
Modeled on higher level

System time scale
Stored in the model

Signal time scale
Seen but forgotten

Noise being filtered
Veiled information

1

constants here only

$$f_c = \frac{1}{2\pi\tau}$$

$$f_c' = \frac{1}{2\pi\tau_x}$$

f

- If the observer has "finite horizon", instead of E{ } one uses

$$\frac{d\mathcal{E}\{z_\iota^2\}}{dt/\tau}(t) = z_\iota^2(t) - \mathcal{E}\{z_\iota^2\}$$

*Lesson: observing real world nonidealities gives rise to emerging structures (here **hierarchies**)*

Semantics – "origin of meaning"

Observer

NOW

Traditionally: interpretation from above

Inheritance from below

World
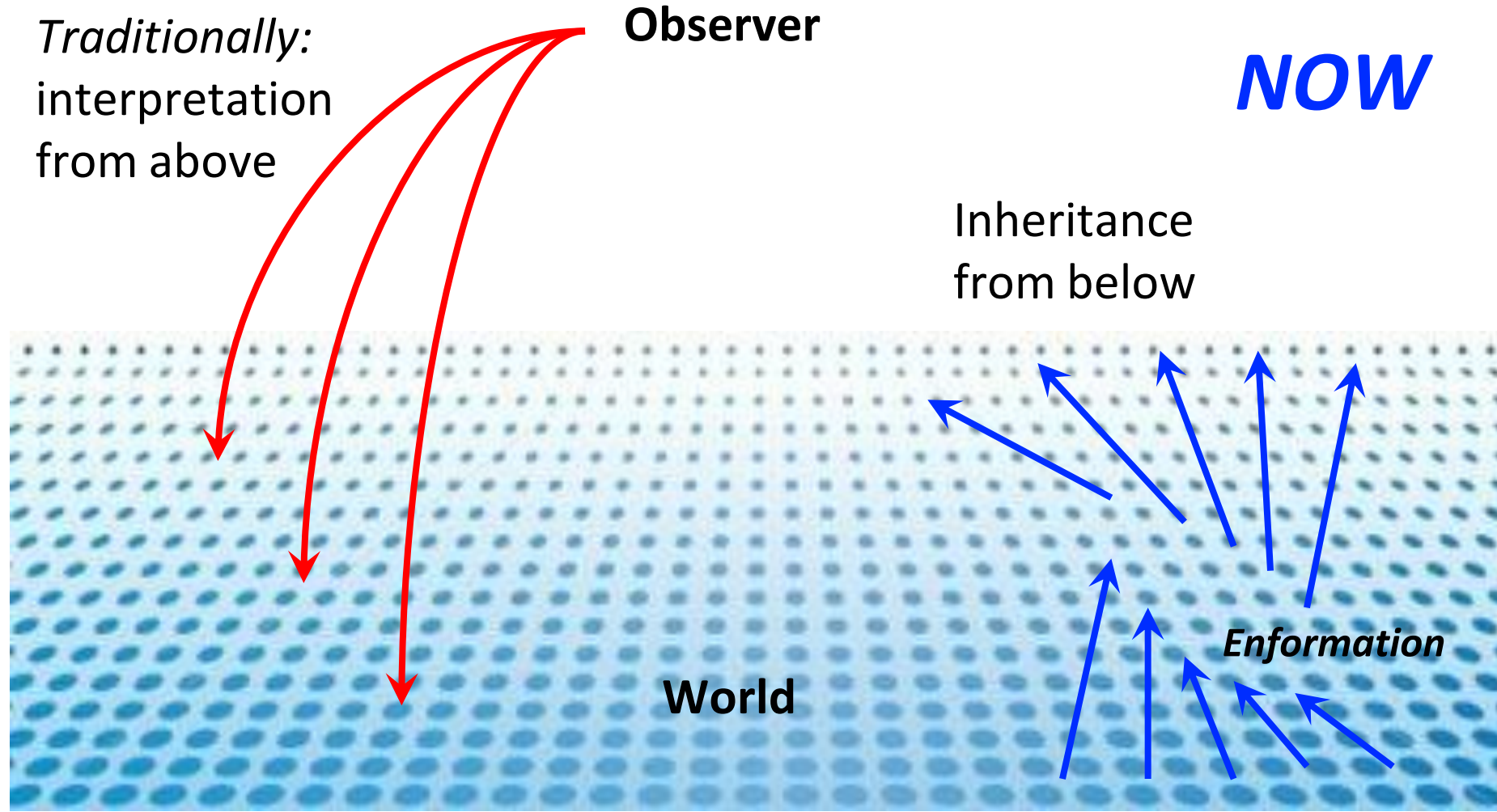
Enformation

Maximize enformation capture = go towards "maximum relevance"!

More general views

- Natural semantics = Cybernetic "Gregory Bateson semantics": everything is based on "differences making difference"
- Bishop Berkeley said that *to exist is to become observed*
- Extending this, one can say that to exist and survive is to *apply* enformation in one's environment ("make one's mark")
- On the other hand, the goal is to *acquire* enformation to be able to exploit it
- Altogether, the goal is *interaction* with the environment
- "Measurement" is the basic functionality in nature: everything measures each other, *man is not needed*
- Optimization: "winner" is the one capturing most enformation; looks goal-directed but this is just an illusion as seen from above

Model of interaction

- Assumption: "Reactions" are a result of "collisions", being proportional to "activities" being related to "concentrations"

$$\frac{d\zeta_i}{dt} = \dot{\zeta}_i = \alpha_i\, z_1^{a_{i1}} \cdots z_m^{a_{im}}$$

- Taking logarithms

$$\log \dot{\zeta}_i = \log \alpha_i + a_{i1} \log z_1 + \cdots + a_{im} \log z_m$$

and differentiating around the nominal point one gets

$$\frac{\Delta\dot{\zeta}_i}{\langle \dot{\zeta}_i \rangle} = a_{i1}\frac{\Delta z_1}{\langle z_1 \rangle} + \cdots + a_{im}\frac{\Delta z_m}{\langle z_m \rangle}$$

Enformation theoretic *system*

- The model becomes linear:

$$\bar{x}_i = a_{i1}\bar{u}_1 + \cdots + a_{im}\bar{u}_m$$

- This can also be interpreted as a model for (truly linear) "generalized diffusion"

- Now there are *inputs* (*resources,* or some kind of *pressures*) and *states* (*activities,* or some kind of *rates of change* or *flows*)

$$\bar{u}_j = \Delta z_j / \langle z_j \rangle$$
$$\bar{x}_i = \Delta \dot{\zeta}_i / \langle \dot{\zeta}_i \rangle$$

Local balance values = "observables"

- A *system* is a set of states sharing the same view of the world

Coordinated micro-level vibration is flow and change on macro-level

Mathematics – augmented reasoning

- Mathematics is needed to extend natural language to seamlessly follow the enformation flow

  - Subsymbolic, fuzzy phenomena beyond the crisp concepts are available
  - High dimensionality and simultaneity can be tackled with
  - Time structures can be modeled and infinity can be attacked
  - Emergence and convergence can be captured
  - Optimization can be carried out, etc.

- Enformation is quadratic – optimal structures will be *linear*

- Normally, mathematics is purely syntactic – now semantics is included in formulations

- Only physically relevant structures are derived?

Regardless of linearity, infinite iteration (see later) restores computational power

Survival strategy of systems

- The acquired internal enformation reveals the evolutionary "fitness" of a systemic mode: If $\bar{x}_i = a_{i1}\bar{u}_1 + \cdots + a_{im}\bar{u}_m$

$$\mathcal{E}\{\bar{x}_i^2\} = a_{i1}\mathcal{E}\{\bar{x}_i\bar{u}_1\} + \cdots + a_{im}\mathcal{E}\{\bar{x}_i\bar{u}_m\}$$

- The maximum strategy can be found applying Lagrangian technique for the constrained problem ($|a_i| = const$), giving

$$a_{ij} = q_i\,\mathcal{E}\{\bar{x}_i\bar{u}_j\}$$

- Also – for all i and j in a system the "surviving" interactions can be formally written as

$$\bar{x} = Q\,\mathcal{E}\{\bar{x}\bar{u}^{\mathrm{T}}\}\,\bar{u}$$

$$\bar{x} = Q \, \mathcal{E}\{\bar{x}\bar{u}^T\} \, \bar{u}$$

- The data structures are

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{pmatrix} \qquad \bar{u} = \begin{pmatrix} \bar{u}_1 \\ \vdots \\ \bar{u}_m \end{pmatrix}$$

"Coupling matrix"

$$Q = \begin{pmatrix} q_1 & & 0 \\ & \ddots & \\ 0 & & q_n \end{pmatrix}$$

"Cross-enformation"

$$\mathcal{E}\{\bar{x}\bar{u}^T\} = \begin{pmatrix} \mathcal{E}\{\bar{x}_1\bar{u}_1\} & \cdots & \mathcal{E}\{\bar{x}_1\bar{u}_m\} \\ \vdots & \ddots & \vdots \\ \mathcal{E}\{\bar{x}_n\bar{u}_1\} & \cdots & \mathcal{E}\{\bar{x}_n\bar{u}_m\} \end{pmatrix}$$

Semantic filter

Everything is still local

Analysis of $\bar{x} = Q\,\mathcal{E}\{\bar{x}\bar{u}^{\mathrm{T}}\}\,\bar{u}$

- If one defines

$$\theta^{\mathrm{T}} = Q^{1/2}\mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}^{-1/2}\mathcal{E}\{\bar{x}\bar{u}^{\mathrm{T}}\}$$

it turns out that

$$I_n = \theta^{\mathrm{T}}\theta$$

$$Q^{-1} = \theta^{\mathrm{T}}\,\mathcal{E}\{\bar{u}\bar{u}^{\mathrm{T}}\}\,\theta$$

- *Principal subspace analysis* gets implemented

- Data covariance structure becomes modified!

"Coupling effect"

Open-loop eigenvalues
Sub-cybernetic system
Marginally cybernetic
"Hyper-cybernetic"

λ_n

λ_{n+1}

Untouched

Unequal stiffnesses

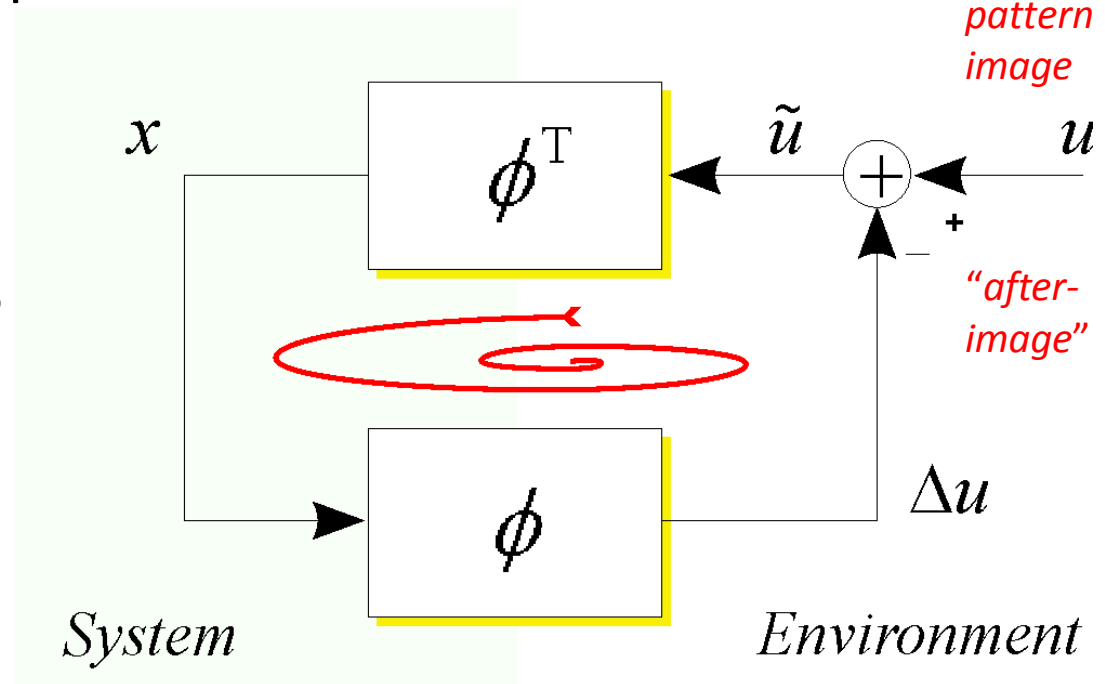$1/q$

$1/q_n$

$1/q_1$

$\bar{\lambda}$

Eigenvalue index
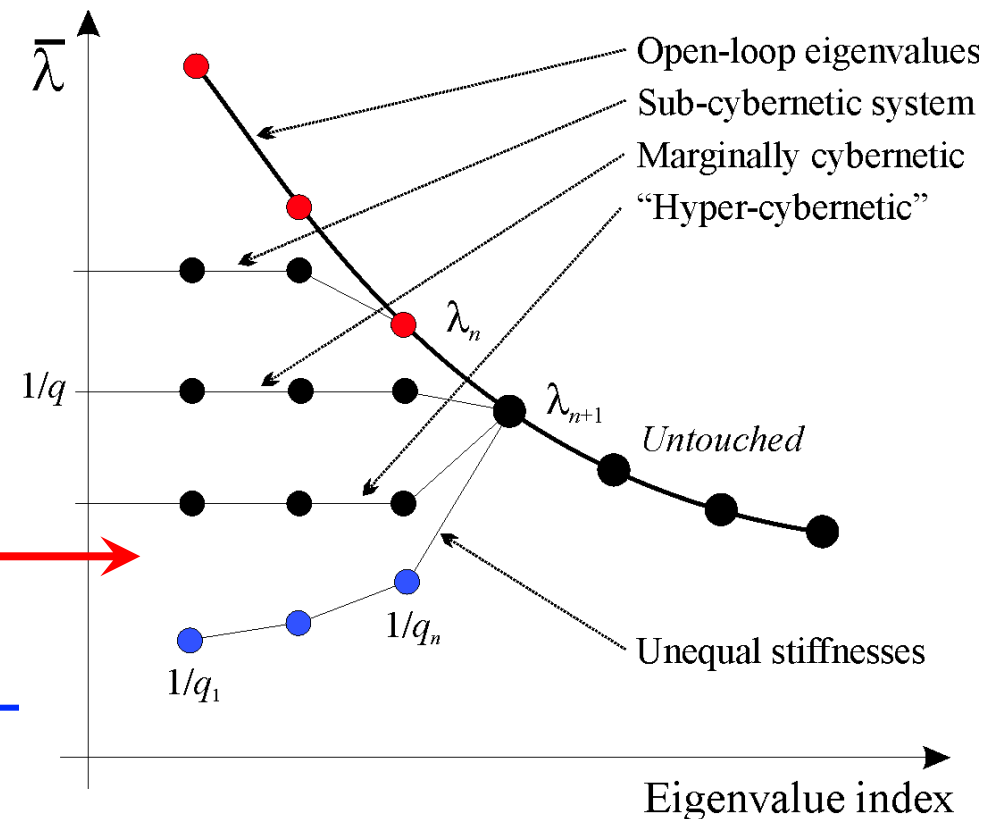
But to make it work…

- The above holds only if a stationary solution can be found
- Here, apply *linear* ("Adam Smith" style) *negative feedback*: enformation is not information, *exploitation means exhaustion*
- Self-organization through competitive adaptation
- "Computational power": Linearity compensated through *infinite iteration*

$$\tilde{u}(t) = u - \Delta u(t)$$

$$\bar{u} = \lim_{t \to \infty} \{\tilde{u}(t)\}$$

x

ϕ^{T}

$\tilde{u}$

$+$

u

pattern image

$+$

$-$

"after-image"

ϕ

Δu

System

Environment

Compare this "stigmergy effect" to traditional Hebbian algorithms

Inclusion of multi-level dynamics

Pandora's simulation box

Statistical level: "Ecolockers" and *diversity*

Unmodeled noise

Cumulated enformation (*structure*)

Sources of enformation (*variations*)

Flow of enformation

Regulation

Cumulated matter (*biomass*)

Flow of matter/energy

Sources of matter/energy (*levels*)

Dissipative waste
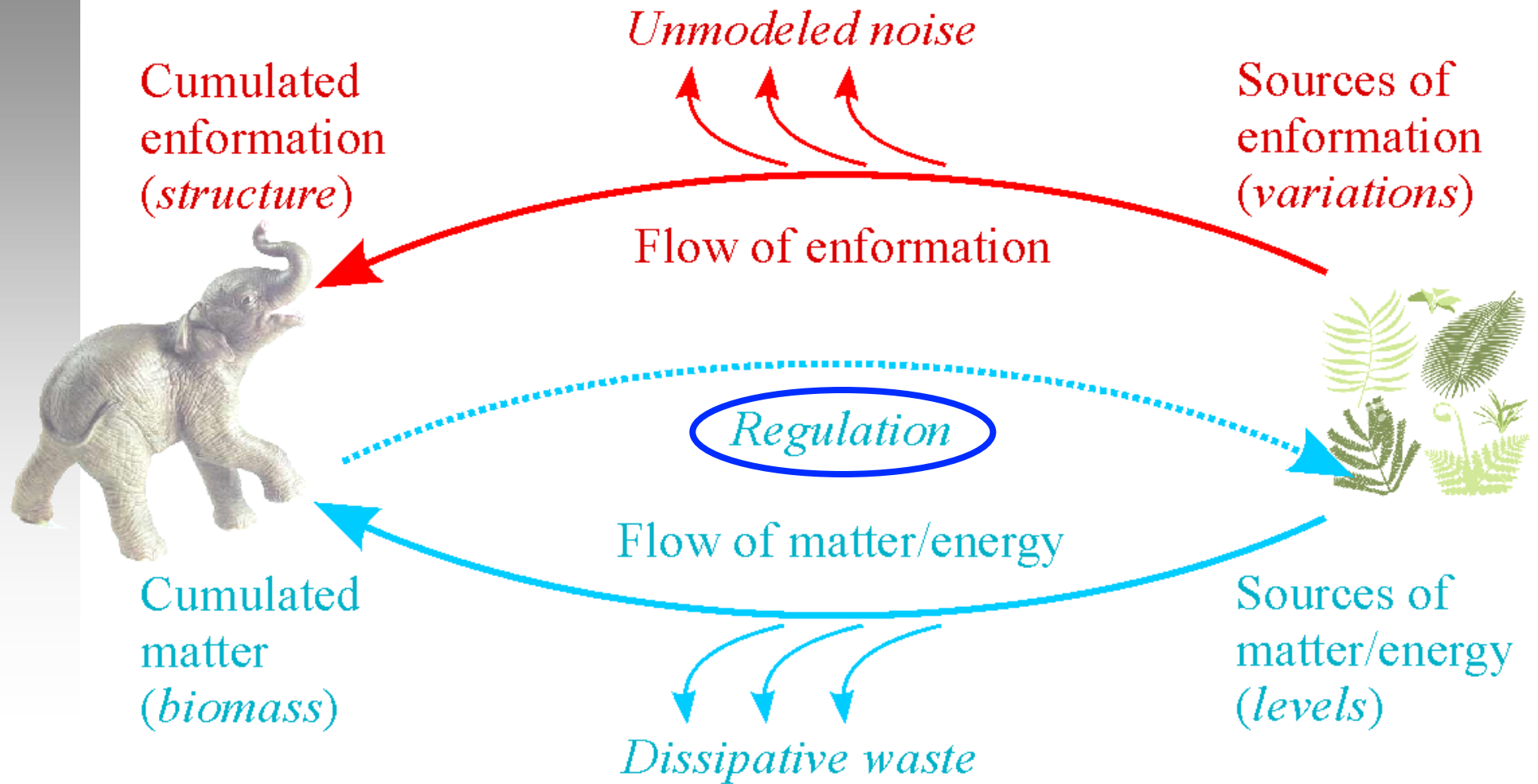
Nonlinearities typically rotate the model towards *sparse components*

Altogether: ~~Optimal~~ control of enformation
robust

- The interaction is governed by the steady-state mappings

$$\phi^{\mathrm{T}} = Q\,\mathcal{E}\{\bar{x}\bar{u}^{\mathrm{T}}\} = \left(Q^{-1} + \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}\right)^{-1}\mathcal{E}\{\bar{x}u^{\mathrm{T}}\}$$

$$\varphi^{\mathrm{T}} = \left(Q^{-1} + \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}\right)^{-1}\mathcal{E}\{\bar{x}\bar{u}^{\mathrm{T}}\}$$

so that in equilibrium

$$\begin{cases} \bar{x} &= \phi^{\mathrm{T}}\bar{u} \\ \bar{x} &= \varphi^{\mathrm{T}}u \\ \hat{u} &= \phi\,\bar{x} \\ \hat{\bar{u}} &= \varphi\,\bar{x} \end{cases}$$

where $\quad \bar{u} = u - \hat{u}$

$\longleftarrow$ Optimal *modeling* of enformation

robust

$\longleftarrow$ ~~Optimal~~ *estimation* of enformation
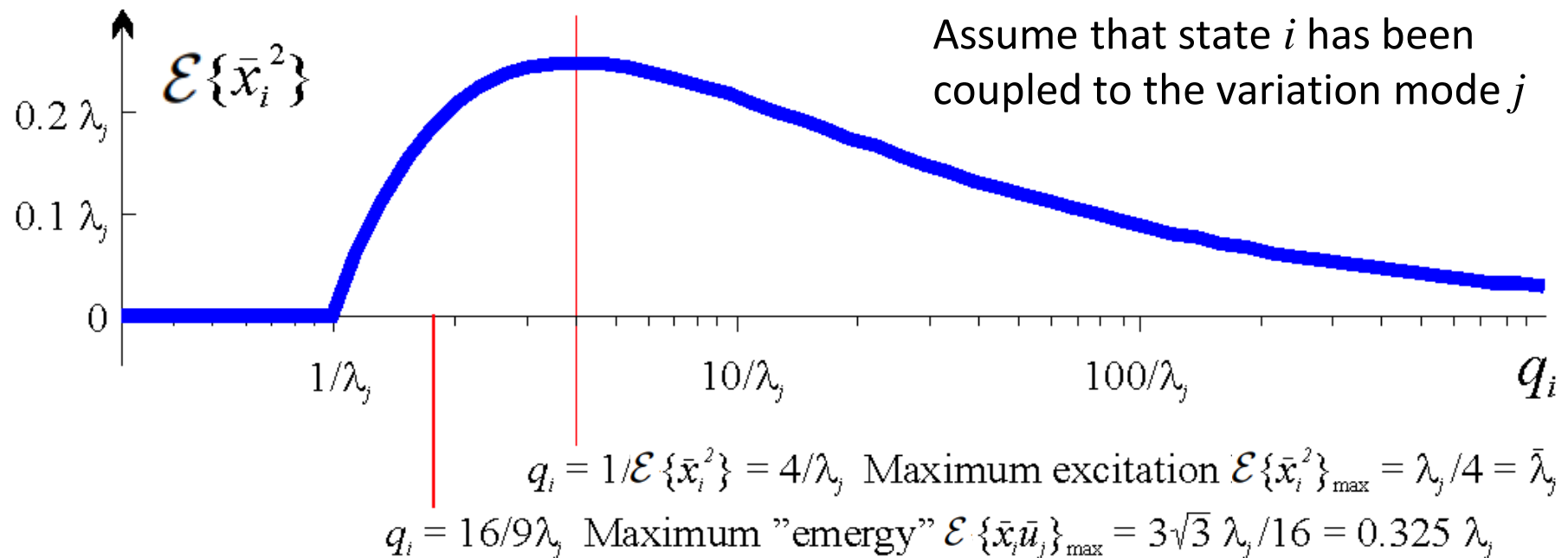
Rather than optimal *least-squares regression*, there is *ridge regression*

"The way up and the way down is the one and the same"

Inheritance of enformation

Assume that state i has been coupled to the variation mode j

$q_i = 1/\mathcal{E}\{\bar{x}_i^2\} = 4/\lambda_j$ Maximum excitation $\mathcal{E}\{\bar{x}_i^2\}_{max} = \lambda_j/4 = \bar{\lambda}_j$

$q_i = 16/9\lambda_j$ Maximum "emergy" $\mathcal{E}\{\bar{x}_i\bar{u}_j\}_{max} = 3\sqrt{3}\,\lambda_j/16 = 0.325\,\lambda_j$

- One has

$$\mathcal{E}\{\bar{x}_i^2\} = \sqrt{\frac{\lambda_j}{q_i} - \frac{1}{q_i}}$$ so that to become coupled $q_i > \dfrac{1}{\lambda_j}$

"Static friction" emerges?

Simplifying the model

- Automatic adjustment of the coupling factor (typically $b = 1$)

$$q_i = b \, \frac{1}{\mathcal{E}\{\bar{x}_i^2\}}$$

  resulting in *maximum excitation* in the system

- Then the system data get seasoned in an interesting way:

  - Variables (system/env modes) become *equalized* = variances become the same
  - The visible environment becomes *whitened* = covariance is a unit matrix
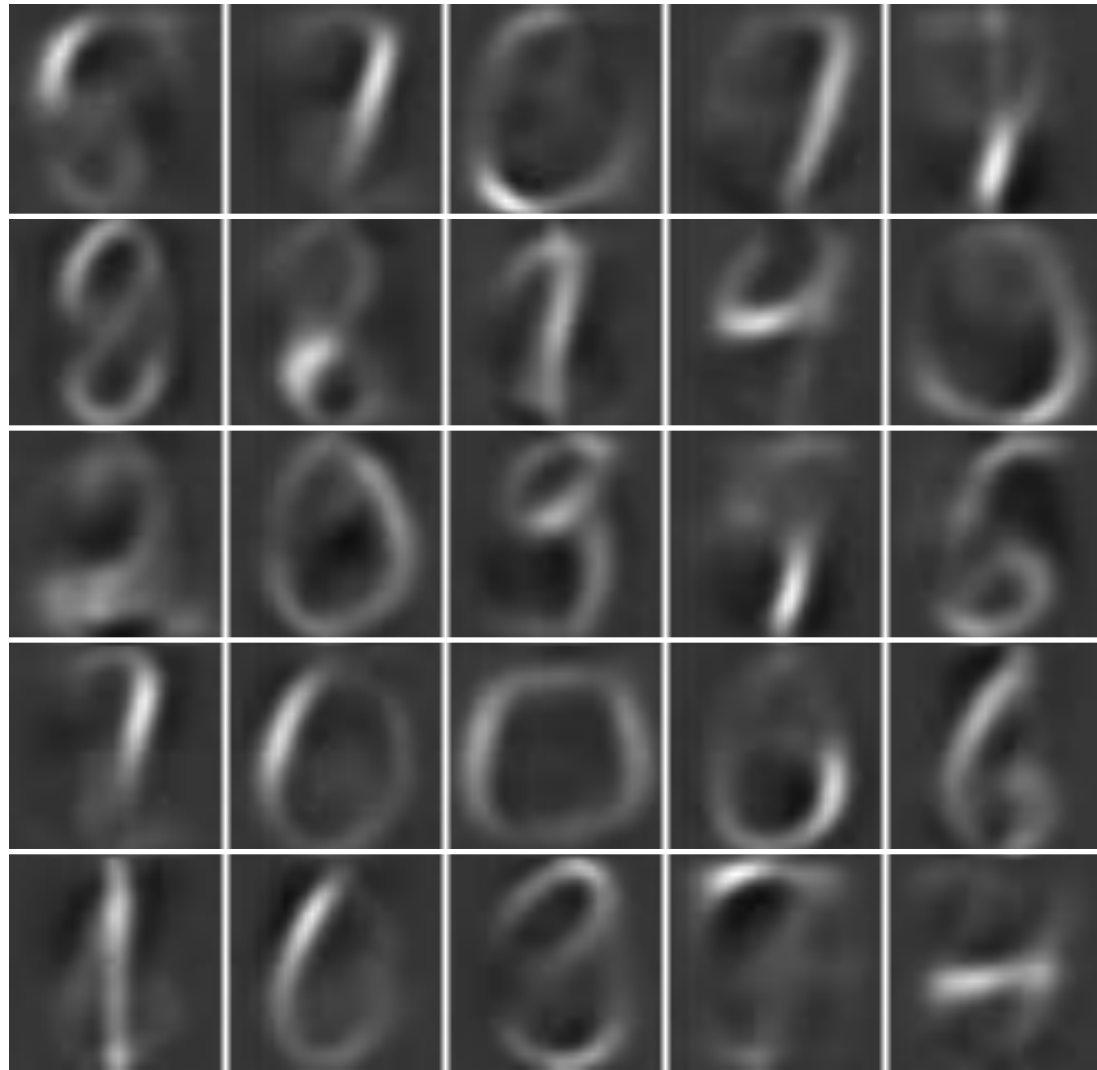
- System size can be optimized according to data properties:

$$\sqrt{\lambda_n} > \frac{1}{2} \frac{\sum_{\iota=1}^{n-1} \sqrt{\lambda_\iota}}{n-1} \qquad \text{must hold for all included modes}$$

Example: *Hand-written digits*

- Features that are learned = 25 vectors ϕ_i

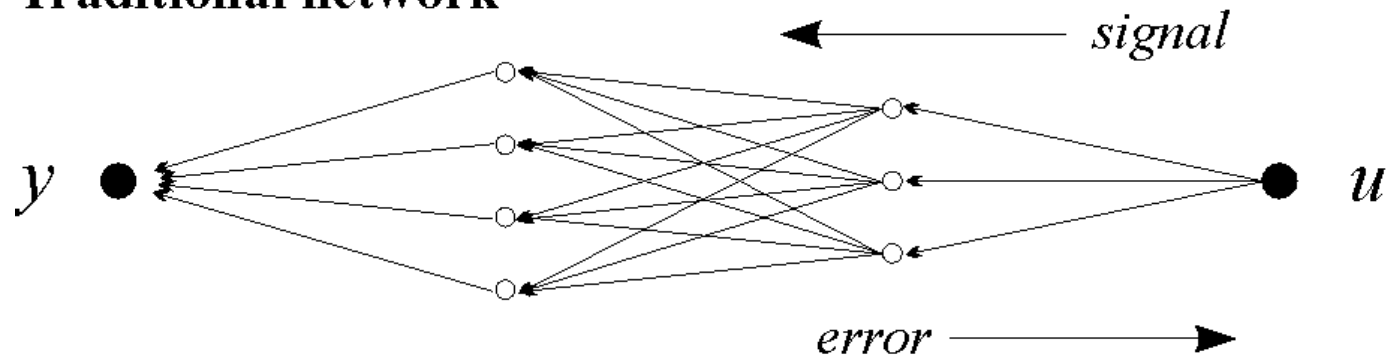- Abs-value nonlinearity ("symmetry breaking")

Epoch 500

Relation to some machine learning practices

- **Hopfield nets**: Similarly, there are *energy functions*, but now stored patterns are decomposed to adaptive features, etc.

- **(Restricted) Boltzmann machines**: Again, signals are filtered in recursion, but now variables are continuous-valued

- **(Generalized/Anti) Hebbian algorithms** and **Oja's rule**: Now structures are linear and physiologically more plausible

- **(Kohonen's) self-organizing maps**: Letting coupling matrix Q be originally non-diagonal, *neighborhood* is implemented

- **Principal/Independent/Sparse Component Analysis {P/I/S}CA**: More or less *straightforward extensions* (?!) of the basic scheme

- **Multilayer perceptrons**: Error signals are now a natural part of the functioning, no *back-propagation* is needed

From "laminar" to "turbulent" flow
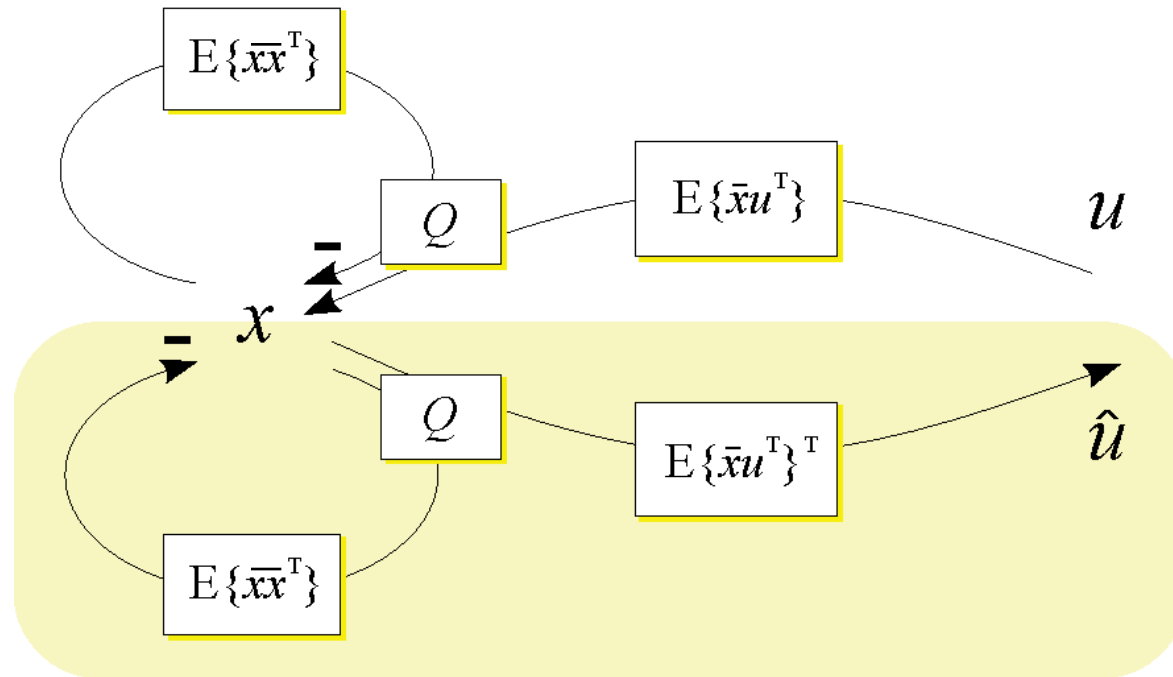

Traditional network

signal

y u

error

Neocybernetic network

enformation

y u

"Whirls" emerging

"neg-enformation"
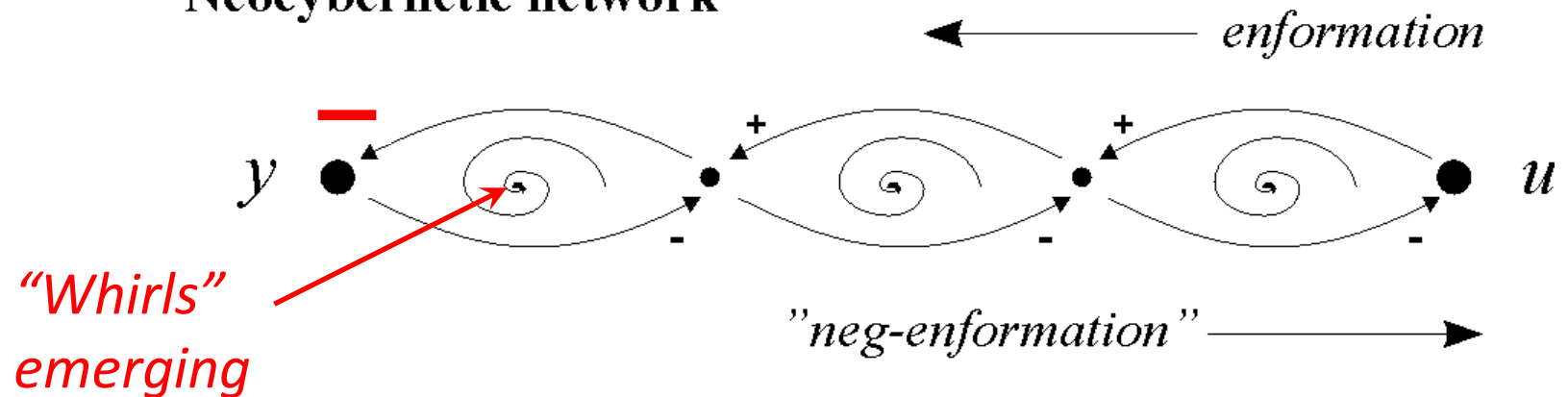
From easy signal propagation and difficult adaptation to complicated signal transfer and easy adaptation

After an evolutionary step

- Context-awareness assumed
- "Anti-Hebbian" part included in model
- Internal lossless feedback, no more dissipation
- Implementation of pure PSA / MLR

$$\bar{x} = \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}^{-1}\mathcal{E}\{\bar{x}u^{\mathrm{T}}\}\,u = \Phi^{\mathrm{T}}u$$

$$\hat{u} = \mathcal{E}\{u\bar{x}^{\mathrm{T}}\}\mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}^{-1}\,\bar{x} = \Phi\,\bar{x}$$

View from above

- Mathematical pattern: cost criterion getting minimized

$$J(x) = \frac{1}{2} x^{\mathrm{T}} \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\} x - x^{\mathrm{T}} \mathcal{E}\{\bar{x}u^{\mathrm{T}}\} u$$

Possible interpretations as *mechanical energy* (applications in "Aristotelian mechanics"), *deformation energy* (structure optimization), *electrostatic energy* (in *molecular orbitals*!?)…

- Pattern matching view:

$$J(x) = \frac{1}{2} (u - \Phi x)^{\mathrm{T}} \Phi \, \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\} \, \Phi^{\mathrm{T}} (u - \Phi x) \quad \textcolor{red}{\textbf{\textit{+ constant}}}$$

Compare this to traditional maximum likelihood matching!
Now weighting towards *freedoms*: $\quad \Phi \, \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\} \, \Phi^{\mathrm{T}} = \mathcal{E}\{\hat{u}\hat{u}^{\mathrm{T}}\}$
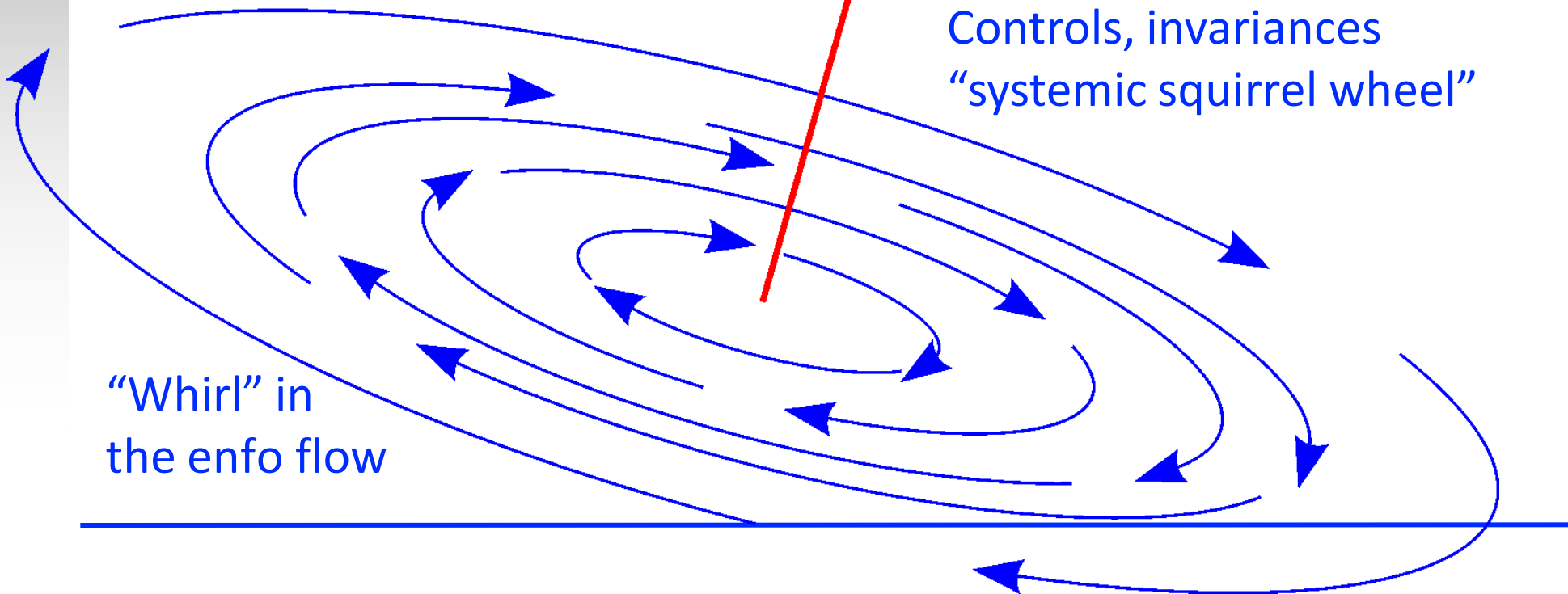
Direction of emphasis ... *and evolution*

- When there are plenty of constraints, it is clever to model only the remaining DOF's (Ockham)

Degree of freedom, direction of covariance, "enformation channel"

Controls, invariances "systemic squirrel wheel"

"Whirl" in the enfo flow

- Above, dynamic thinking was the key to new functionalities – now, study the details of finding the asymptotic state

- There is again a wealth of possibilities – still, the guiding principle of optimality is available

- Expectation Maximization approach applying Ensemble Kalman filter

$$x^{\text{posteriori}} = x^{\text{priori}} +$$

$$C\Phi^{\text{T}} \left(\Phi C\Phi^{\text{T}} + R\right)^{-1} \left(u - \Phi x^{\text{priori}}\right)$$

gives:

$$\frac{dx}{dt/\tau_x}(t) = G\,\Phi^{\text{T}} \left(u - \Phi x(t)\right) = G\,\Phi^{\text{T}}\,\tilde{u}(t)$$

Towards *frequency* (*Laplace, Fourier*) *domain*

- When the system "input" is another dynamic entity

$$\frac{d\,\tilde{u}}{dt/\tau_u}(t) = -\gamma\,\Phi G^{\mathrm{T}}\,x(t)$$

- When combined, the overall dynamics becomes

$$\frac{d^2 x}{dt^2/\tau_x\tau_u}(t) = G\Phi^{\mathrm{T}}\frac{d\,\tilde{u}}{dt/\tau_u}(t) = -\gamma\,GG^{\mathrm{T}}\,x(t)$$

resulting in *harmonic oscillations*:

$$x(t) = A\sin\left(\sqrt{\frac{\gamma}{\tau_x\tau_u}GG^{\mathrm{T}}}\,t + \psi\right)$$

Wrong in the book!

Another qualitative leap in expressional power

- On the next level: Oscillating signals constitute *fields*, systems being characterized through interacting *spectra*
- Freedom of absolute place and time – instead, *amplitudes* and *phases* become relevant (enformation interpretation is still valid)
- Enformation transfer necessitates *impedance matching* among systems, etc.
- Systems constitute *directed antennas*, making it possible to understand *emergent physical structures*
- Coupling to the environment means *collapse of fields*, so that resulting *standing waves* match the boundary conditions
- Applies also to *consciousness* (?!) – and *intersubjective* coupling means that *morphic fields* need to be *simulated*…

A mathematics wizard would be needed …!

Final catch: *information* vs. *enformation*

- Chaos (possibility of models) changes to *structure* and *uncorrelated noise*, both having high (Shannon) information content
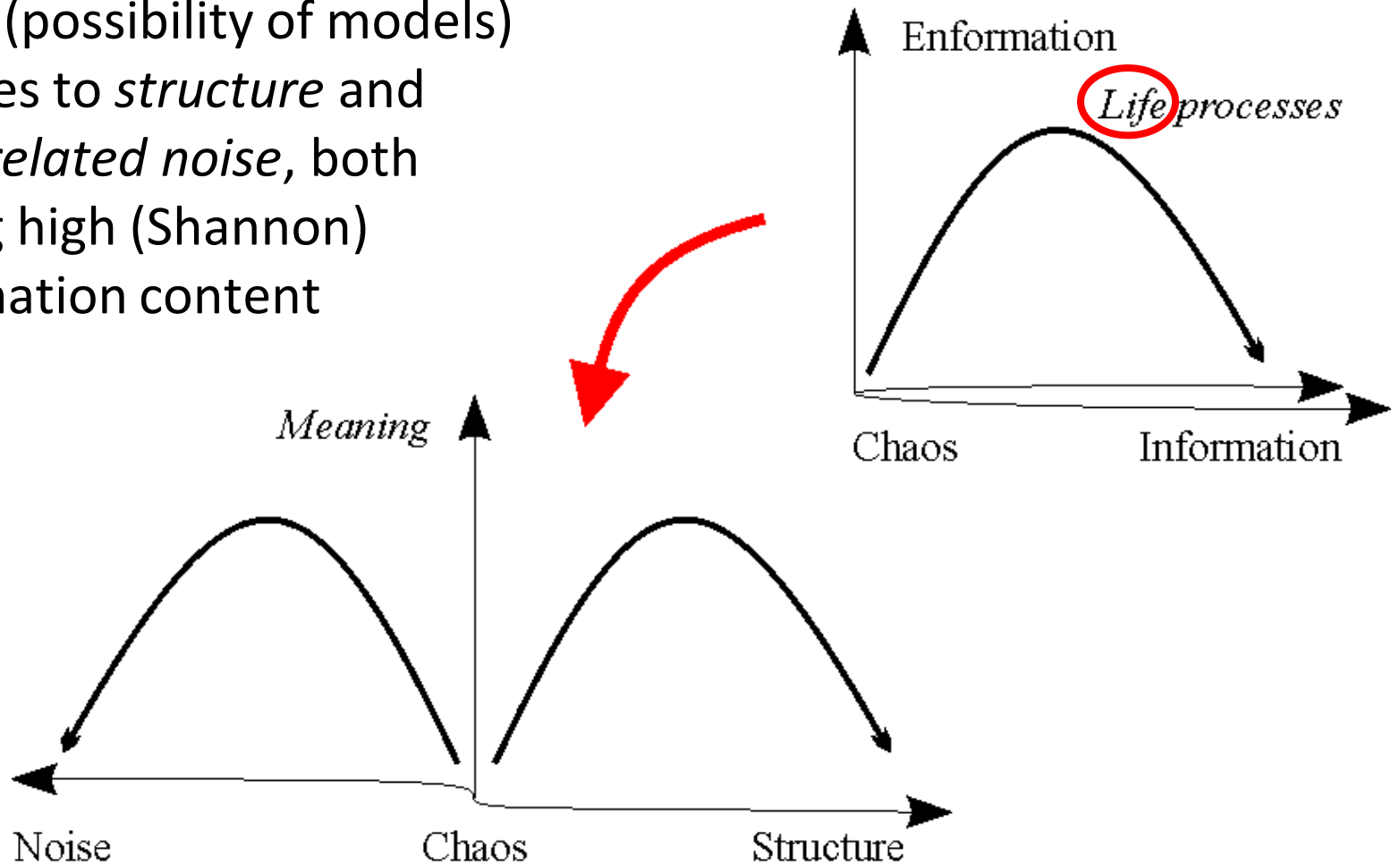
Nature of evolution

structure

system

+ dissipative noise

Fresh enformation is essential – it changes to rigid control structures

chaos = possibility of model
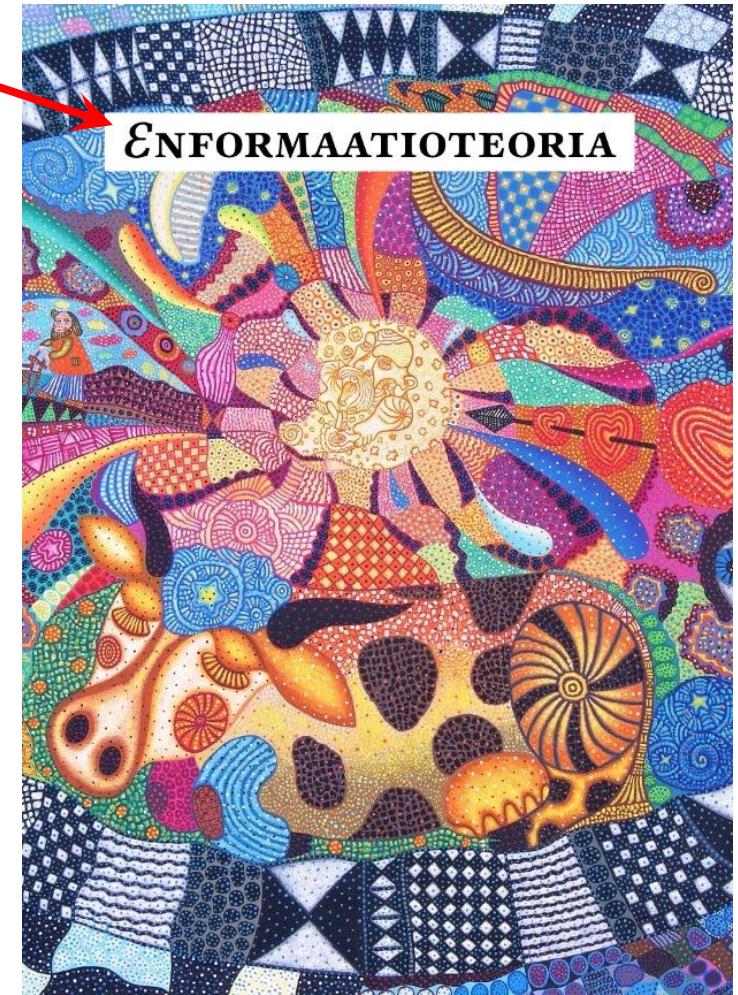
Boundary between order and chaos
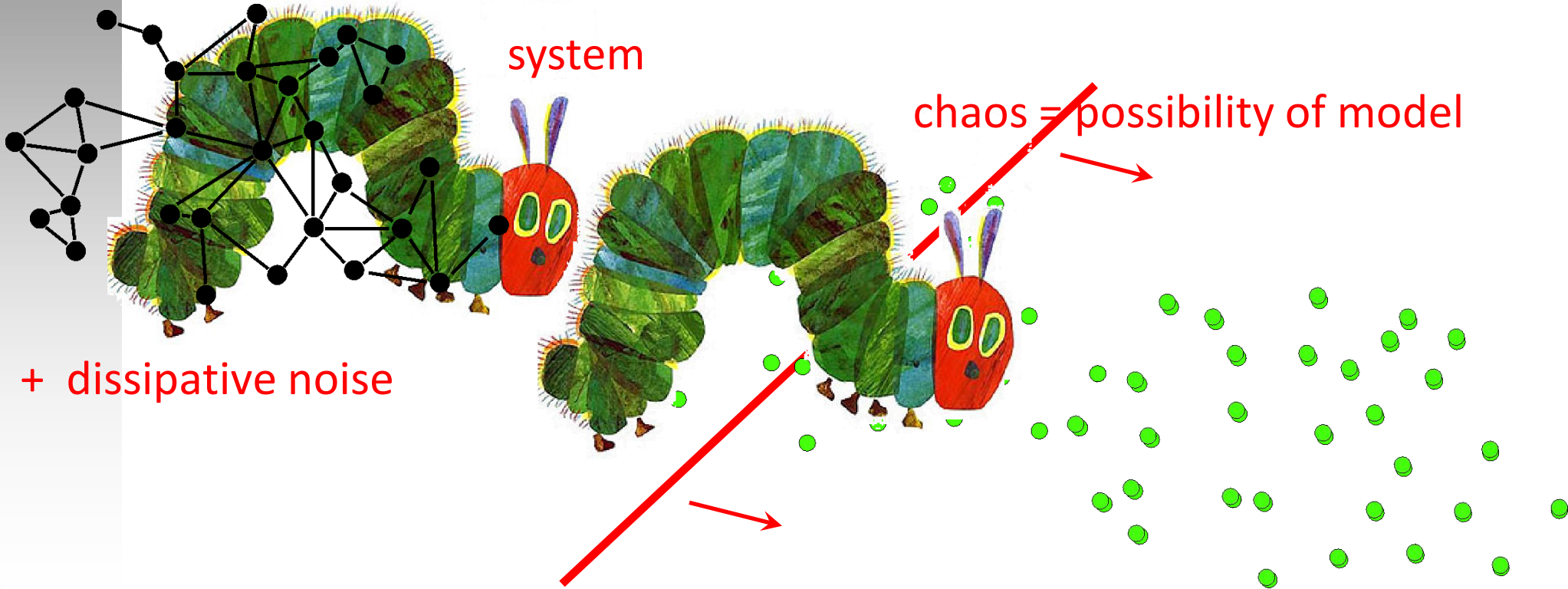
Life, Universe, and $\mathcal{E}$*verything*

- General theory of *life* (in Finnish)

- Life = Fractal structure of "whirls" (control loops, *enformation pumps*) in enformation flow

- *Entropy production is maximized* – but *only with regard to a model*

- In the *subjective world* only there is possibility to *consistency…*

We are happy to explain!

Feel free to contact.

ENFORMAATIOTEORIA