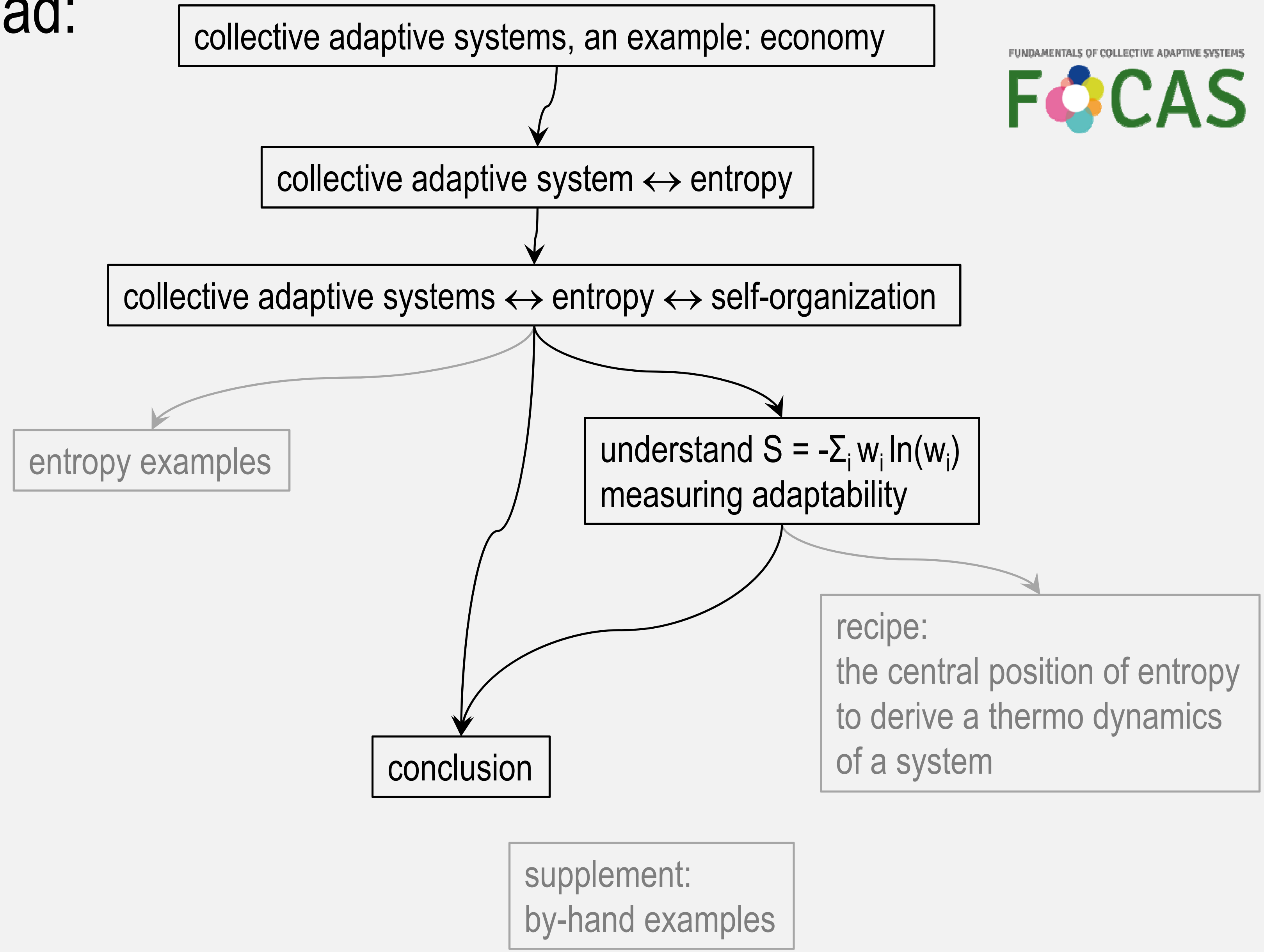
ECAL 2013

FUNDAMENTALS OF COLLECTIVE ADAPTIVE SYSTEMS
FOCAS

Can Adaptability be measured?
Yes, it's Entropy!

Rupert Reiger
rreiger@online.de
rupert.reiger@eads.net

thread:


collective adaptive systems, an example: economy
collective adaptive system ↔ entropy
collective adaptive systems ↔ entropy ↔ self-organization
entropy examples
understand $S = -\Sigma_i w_i \ln(w_i)$
measuring adaptability
recipe:
the central position of entropy
to derive a thermo dynamics
of a system
conclusion
supplement:
by-hand examples

collective adaptive systems ↔ economy

- This is to interrelate the subject of
  - economy, markets, and indicators thereof e.g.
    - innovation / patents
    - distribution of company seizes: Big → SME
    - number of start-ups
  - to collective adaptive systems (CAS)

  - economy is a good example for a CAS
    - with its ingredients of employments, markets, humans using computers
    - generating innovation and adaptation on different system scales: personally, on company scales, nationally
    - it is a network of humans and means, optimizing locally to a system-total optimum in a chaotic, non-equilibrium system

collective adaptive systems ↔ economy

- CASs differ from current generation systems in two aspects:
  - collectiveness: Systems typically comprise multiple units
    - which have autonomy in terms of their own individual properties, objectives and actions
    - decision-making is highly dispersed and interaction between the units may lead to the emergence of unexpected phenomena and behaviors
  - multiple scales in time and space
    - systems are open, in that units may enter or leave at any time
    - the units can be operating at different temporal and spatial scales and
    - have competition → different potentially conflicting objectives and goals on any level
  - economy is fulfilling all of these criteria

excerpting from "Collective Adaptive Systems Expert Consultation Workshop 3 & 4 November 2009"

collective adaptive systems ↔ economy

- CAS constitute a new kind of physical intelligence, a well-known fact in economy
  - known as the "invisible hand of the market" by Adam Smith (1723-1790) to describe the self-regulating behavior of the marketplace and
  - as "Competition as a Discovery Procedure" Friedrich Hayek's (1899-1992) as a self-organization procedure. Hayek is addressing, that local adaptation, evolution and competition is the process to discover innovation and wealth
- scaling properties of CAS are easy to be seen in economy by zooming in
  - from autonomously acting computers → autonomously acting single persons → autonomously acting divisions → autonomously acting companies → autonomously acting national economies → different competing global economical systems
  - by acting, making patents, working, making decisions, optimizing on every level in a scale invariant manner
- physically speaking, economy is chaotic in principal
  - economic indicators like stock indexes are sensitive to their start values and diverging exponentially by their Lyapunov exponent
  - averaging this out by applying some other models like game theory must never lead to the right future results, chaos can not be averaged
  - these systems can be fitted to the past, but
  - although deterministic, future development is not predictable in principle

F.A. Hayek, *Competition as a Discovery Procedure*, e.g. THE QUARTERLY JOURNAL OF AUSTRIAN ECONOMICS VOL. 5, NO. 3 (FALL 2002): 9–23: https://itunesu.mises.org/journals/qjae/pdf/qjae5_3_3.pdf

collective adaptive systems ↔ economy

- Evolutionary computing is not the same as biological evolution
  - in nature diversity (entropy) tends to increase through the evolutionary tree, whereas in computing it decreases
  - in fact, both are always present and economy is best to study that:
    - the decrease of diversity in economy is resulting from the little big global winners and the extinction of the competitors (towards low entropy)
    - the increase of diversity is resulting from the big global winners over-organization resulting in rigidness in innovation so having the bigger probability for innovation to happen outside the big winners (coming from outside high entropy)
    - entropy is delivering success from the bottom, ideas from the basis that established big players did not have: So IBM missed Microsoft, Microsoft missed Google, Google missed Twitter, Twitter missed Facebook
  - physically speaking, the system is far from equilibrium condition
  - in a closed system entropy is fixed, or increasing in the presence of irreversible processes = 2nd law of thermodynamics
  - in an open and/or non-equilibrium system, entropy can locally decrease with entropy export
  - entropy is often said to be a measure of disorder, that is wrong
  - entropy is a measure of diversity and a measure of information which can be gained in case of an event

excerpting from "Collective Adaptive Systems Expert Consultation Workshop 3 & 4 November 2009"

collective adaptive systems ↔ entropy
macrostates ↔ microstates

- a physical example is a simple gas of mass points with
  - a macrostate described by: temperature, pressure, inner energy, volume
  - “one” microstate of N point-particles is described by the 6N variables of every points location and momentum fulfilling the macrostate
  - there are a lot of these and this number has to be maximized, for:
- Entropy is a measure of in “how many ways” a condition or some macrostate can be fulfilled
- “Each of these ways” specifies a microstate
- The macrostate with the larger number of microstates is the more probable one
- The most probable state is the one which maximizes the number of microstates
- That is the equilibrium state

collective adaptive systems ↔ entropy

ECAL 2013

FUNDAMENTALS OF COLLECTIVE ADAPTIVE SYSTEMS
FOCAS

- Entropy interpretations:
  - entropy = 0 → all in one microstate → full predictability
  - That is Nernst's theorem = 3rd law of thermodynamics
  - higher entropy →
    - higher disorder, that is not a good description but:
    - the more possibilities for microstates we have (think about each one of an entity capable of innovation)
    - the less the degree of predictability
    - the less our knowledge before an event happens
    - the higher the gain of information in case of an event
- For economical systems:
  - the higher the entropy, the higher the capability for innovation !
  - that's the core for a self-organizing system together with adaptability, and
  - that is lowering entropy !
  - **these effects are opposing**

collective adaptive systems ↔ entropy ↔ self-organization

- That seems to be a paradox, for
  - an isolated system is keeping or raising entropy
  - a system with maximal entropy cannot change on its own, but
  - when a chemical reaction is strongly exothermic, then the entropy inside the reactive system can be lowered; but that is at least compensated by the freed energy, raising the entropy in the surroundings
  - but also in an open system:
  - lowering entropy by self-organization is based on importing energy and exporting entropy in any way (different from being exothermic), resulting in a raise of entropy in sum
  - energy import from sun to earth and entropy export from earth to space is the precondition for self-organization, for the emergence of live on earth
  - so complex structures can arise by self-organization not contradicting the 2nd law of thermodynamics

e.g.: Komplexe Strukturen: *Entropie und Information*, Jan Freund, Frank Schweitzer, Werner Ebeling, ISBN-13: 978-3815430323

entropy

number of all microscopic states the system can have

Boltzmann's grave at the Zentralfriedhof Vienna, showing S = k log W

entropy examples

ν counts the microstates

fundamental equation

$$S = \max(-k \sum_{\nu} w_{\nu} \ln w_{\nu})$$

k Boltzmann constant

w_{ν} probability for an event

x counts N objects in a bin

fundamental equations information theory

$$S = -\sum_{x} p(x) \log_2 p(x)$$

or also :

$$S = - \lim_{N \to \infty} \sum_{i} \left(\frac{n_i}{N}\right) \ln\left(\frac{n_i}{N}\right)$$

or also :

$$S = \frac{1}{N} \ln W = \frac{1}{N} \ln \frac{N!}{\Pi_i n_i!}$$

microcanonical ensemble

$$S = k \ln W$$

with :

all microstates have the same energy U

W is the number of all microstates with energy U

canonical ensemble

energy can enter and leave the system

$$S = k \ln Z + \frac{1}{T} U$$

with :

$$Z = \sum_{\nu} e^{-U_{\nu}/kT}$$

Z partition function

grand canonical ensemble

energy and particles can enter and leave the system

$$S = k \ln Z + \frac{1}{T} U - \frac{\mu}{T} N$$

with:

$$Z = \sum_{\nu} e^{-U_{\nu}/kT + \mu N_{\nu}/kT}$$

Z partition function

μ chemical potential

entropy examples

fundamental equation

$$S = \max(-k \sum_{\nu} w_{\nu} \ln w_{\nu})$$

k Boltzmann constant

w_{ν} probability for an event

in Bose and Fermi statistics
that is correct anyway

n_i is the number of particles in a state
so it counts objects in a bin,
upper ν counts various break-downs
of that objects in a bin

canonical ensemble

particles indistinguishable and for high T

$$S = k \ln Z + \frac{1}{T} U$$

$$\frac{N!}{n_1^{\nu}! n_2^{\nu}! \ldots n_m^{\nu}!} = 1$$

that makes the problem not not anymore separable,
but for a high T approximation:

$$Z = \frac{1}{N!} z^n$$

with

$$z = \sum_i e^{-\frac{\varepsilon_i}{kT}}$$

corrected Boltzmann

wrong !

n_i is the number of particles in a state
so it counts objects in a bin,
upper ν counts various break-downs
of that objects in a bin

canonical ensemble

particles distinguishable

$$S = k \ln Z + \frac{1}{T} U$$

$$Z = \sum_{\nu} \frac{N!}{n_1^{\nu}! n_2^{\nu}! \ldots n_m^{\nu}!} e^{-\frac{1}{KT} \sum_{i=1}^{m} n_i^{\nu} \varepsilon_i}$$

or

$$Z = z^n$$

with

$$z = \sum_i e^{-\frac{\varepsilon_i}{kT}}$$

Boltzmann

x counts N objects in a bin

fundamental equations
information theory

$$S = -\sum_x p(x) \log_2 p(x)$$

or also :

$$S = -\lim_{N \to \infty} \sum_i \left(\frac{n_i}{N}\right) \ln\left(\frac{n_i}{N}\right)$$

or also :

$$S = \frac{1}{N} \ln W = \frac{1}{N} \ln \frac{N!}{\Pi_i n_i!}$$

=

≠

to understand $S = -\Sigma_i w_i \ln(w_i)$

$S = k \log W$ and $S = -\Sigma_i w_i \ln(w_i)$ are equal

Look at events i (i = 1, 2, 3, … ,n) occurring with probabilities w_i, $\Sigma_i w_i = 1$ and $0 \leq w_i \leq 1$

What information we get when an event i occurs?

- If an event occurs with absolute certainty, so $w_i = 1$, we do not gain any information if it occurs, so we take information $I(w_i = 1) = I(1) = 0$
- The less probable an event is, the more information we gain when it occurs. So $I(w_i)$ is to be monotonous with $1/w_i$
- For coupled events we have probability $w_i w_j$ and for that we ask additivity, so $I(w_i w_j) = I(w_i) + I(w_j)$

That leads us to: $I(w_i) = -\text{const} \ln(w_i)$

to understand $S = -\Sigma_i w_i \ln(w_i)$

ECAL 2013

FUNDAMENTALS OF COLLECTIVE ADAPTIVE SYSTEMS

FOCAS

- Before an event occurs, we do not know which event will occur and so we will not know which information we will gain $\rightarrow$ there is no help for that, no way
- But what we can calculate is the "**mean of the expected information**" to occur:

  $S' = \Sigma_i\ w_i\ I(w_i) = -\ \Sigma_i \text{ const } w_i \ln(w_i)$
- It is as bigger, as less the occurrence of events i is fixed by big or small w_i and
- vanishes when one event "occurs" with absolute certainty (w_i=1) and so all the others are "not occurring" with absolute certainty ($w_j = 0$, $i \neq j$) since $\lim_{w_j \rightarrow 0} w_j \ln w_j = 0$
- That is the situation, when a system is at "the one ground state at T=0"
- That is Nernst's theorem = 3rd law of thermodynamics that the entropy vanishes at T=0
- We call that mean of to be expected information also the degree of uncertainty, which is measuring how big our unaware before occurrence of an event is, or how much information we gain in case of the occurrence of an event

to understand $S = -\Sigma_i w_i \ln(w_i)$

- Is w_i the probability to measure a certain single state i, then we set the maximum of the degree of uncertainty $S' = -\text{const}\ \Sigma_i w_i \ln(w_i)$ equal to the Entropy of the system: $S = \max(-\text{const}\ \Sigma_i w_i \ln(w_i))$
- This maximization determines the w_i if they are not measurable directly but only the S of the state

to understand $S = -\Sigma_i w_i \ln(w_i)$

- Look at the probabilities w_i, as the probabilities of
  - having innovative ideas, good patents
  - finding companies in certain business volume slots from SME to big
  - finding shares nationally traded also sorted by company volume
  - ability of genes to change and to adopt a system to an environment
  - something else measureable from any system capable to adopt to its environment for measureable success
- We know already that the entropy for these events
  - is lower, the more the probability w_i for that events is to find only in a few peaks, and
  - it's higher, the more these events are broadly distributed, showing that these changes e.g. innovations, genes flips can occur everywhere any time

to understand $S = -\Sigma_i w_i \ln(w_i)$

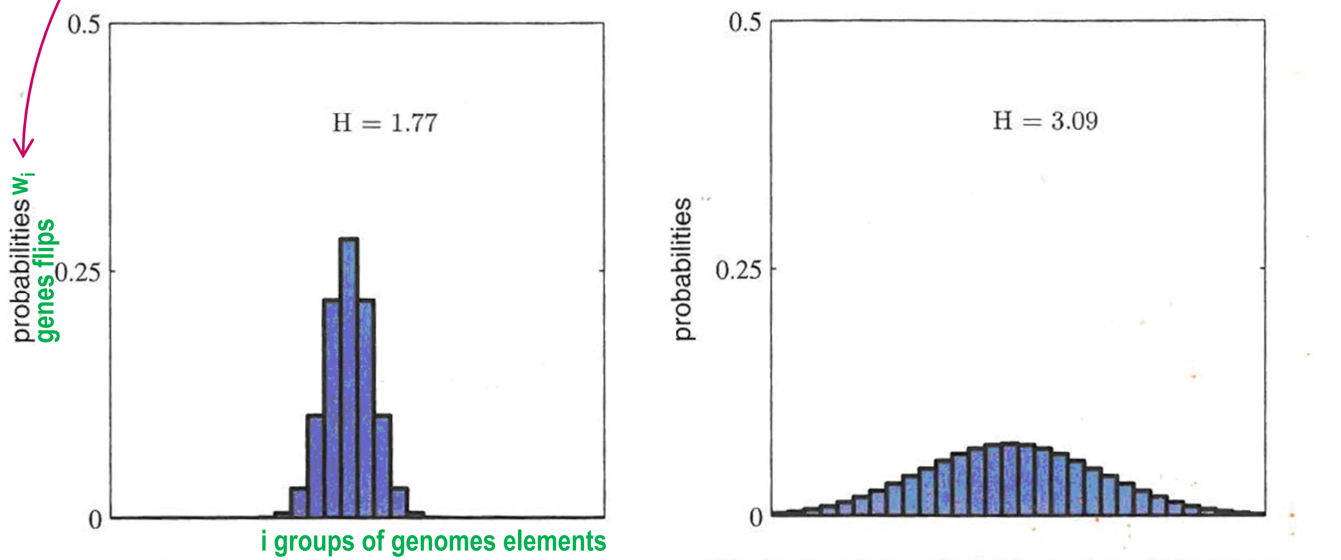
$S = k \log W$ and $S = -\Sigma_i w_i \ln(w_i)$ are equal

Let the states be the entities where innovation can happen, have the extreme values:

- ❖lowest entropy → maximum order → all contents in one state → we know all outcome before →
- ❖→ when states are where innovation can happen, we have only one possibility→ innovation implausible

- ❖highest entropy → equal distribution over all states → maximum uncertainty of the outcome →
- ❖→ innovation can happen everywhere → maximal probability for innovation

The figure illustrates examples inside these extreme values:



Histograms of two probability distributions over 30 bins illustrating the higher value of the entropy H for the broader distribution. The largest entropy would arise from a uniform distribution that would give $H = -\ln(1/30) = 3.40$.

Christopher M. Bishop - Pattern Recognition and Machine Learning
http://www.amazon.de/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738/ref=sr_1_1?ie=UTF8&qid=1340194027&sr=8-1

to understand $S = -\Sigma_i (n_i/N) \ln(n_i/N)$

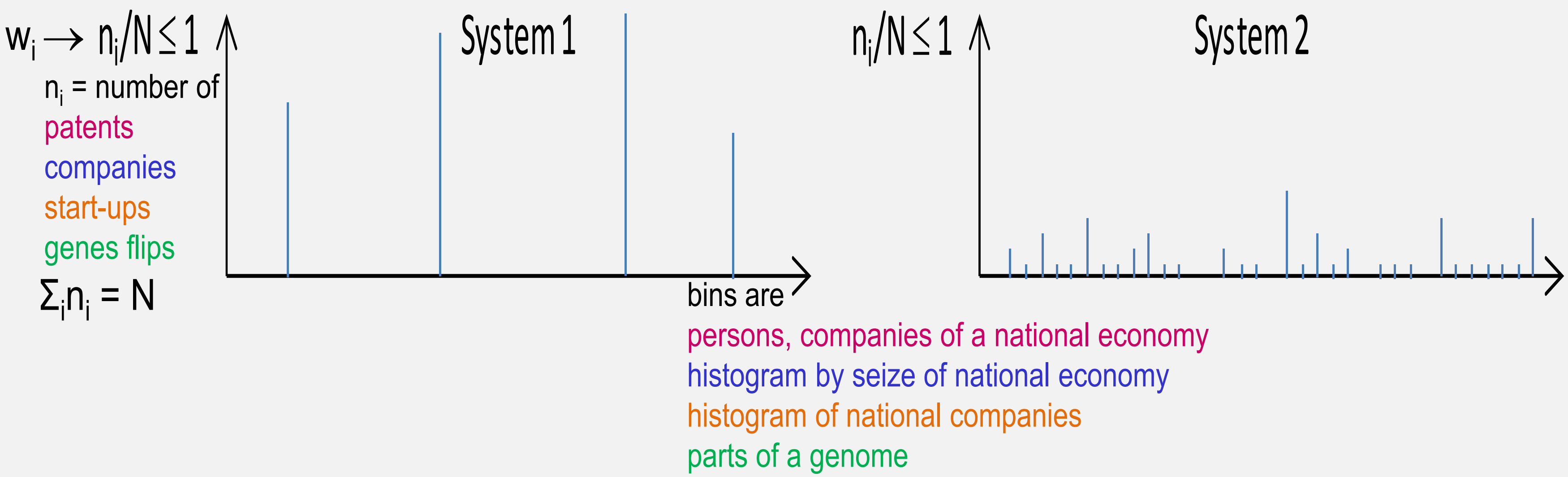
- A simplest approach can be a form of innovation or adaptability entropy:
- $S = -\Sigma_i w_i \ln(w_i) \approx -\Sigma_i (n_i/N) \ln(n_i/N) \quad w_i \leq 1 \quad \Sigma_i n_i = N \quad w_i = \lim_{N\to\infty} (n_i/N)$
- n_i is the number of e.g. patents per entities i like
  - individuals or
  - companies or
  - institutes, or
  - any granularity of some national economy or whatever is wanted to sum over
- N is the total of all of that
- Setting $S = -\Sigma_i (n_i/N) \ln(n_i/N)$ for limited N means in other words using only one existing microstate we know
- The result is one single number, the innovations entropy or adaptability of that national economy, or what other scope is defined
- Doing the calculations on a company level allows comparing the entropy and so innovativeness inside big companies to the ones outside big companies

to understand $S = -\Sigma_i (n_i/N) \ln(n_i/N)$

ECAL 2013

FUNDAMENTALS OF COLLECTIVE ADAPTIVE SYSTEMS FOCAS

- The Figure illustrates examples inside between the extreme values
- System 2 has higher Entropy as System 1
- This could be counts of patents, per person, per enterprise, per organization of a nation
- It is to mention, that entropy is an extensive thermodynamic variable, so doubling the systems means doubling the entropy, good for China, good for a big gene pool.



recipe: the central position of entropy to derive a thermo dynamics

- deriving all dynamics from a statistical beginning:
  - determine all variables (e.g. N,V,E) for the macro state of the system
  - maximize fundamental $S = - k \Sigma_\nu \; w_\nu \ln w_\nu$ to get S and the probabilities w_ν to find microstate ν in the macro state
  - or directly use $S = -\Sigma_i (n_i/N) \ln(n_i/N)$, without maximization the system must not be in equilibrium
  - alternatively:
  - determine the number of all possible microstates (W(N,V,E)) accessible to the system
  - then, the entropy follows from the fundamental S(N,V,E) = k ln W(N,V,E)

  - then it follows further:

  - $P = -(\delta E/\delta V)_{N,S}$ $\mu = (\delta E/\delta N)_{V,S}$ $T = (\delta E/\delta S)_{N,V}$
  - $A = E-TS$ free energy
  - $G = A + PV = \mu N$ Gibbs free energy
  - $H = E + PV = G + TS$ enthalpy

conclusion

- Local competition is the process to discover innovation, knowledge and leads to self-organization
- Spontaneous self organization is, like any organizing principle, lowering entropy and so an opposing trend to the enhancing entropy principle (2nd law) and so also to innovation
- The discovery mechanism is about the discovery of knowledge which would never be produced by any mechanism for their intrinsic chaotic-theoretical manner
- Nothing can be predicted, but it can be measured how an economic society is prepared to let innovation happen by
- one number: entropy

conclusion

- The mechanism of self organization and the mechanism of high entropy are driving a dynamical system in opposite directions
- This typically happens on different scales, having
  - self organization inside companies, lowering entropy and
  - so entropy must be enhanced outside companies
  - so innovation typically happens on a national level outside big companies, so
  - for them it's left to buy the startups, they have no other chance

conclusion

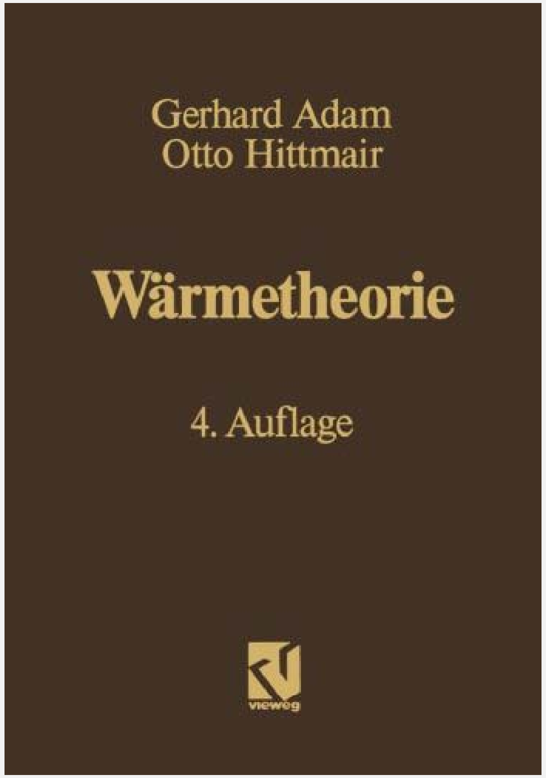
- Entropy is an extensive quantity, so doubling the system doubles the entropy so:
- Size matters (good for China)

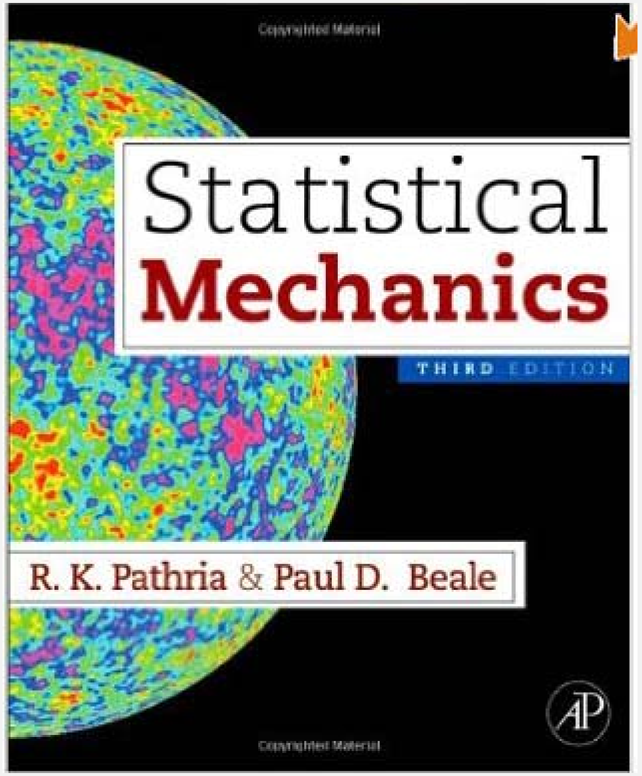
- But:
- Taking e.g. “entropy / national population” shows optimization potential

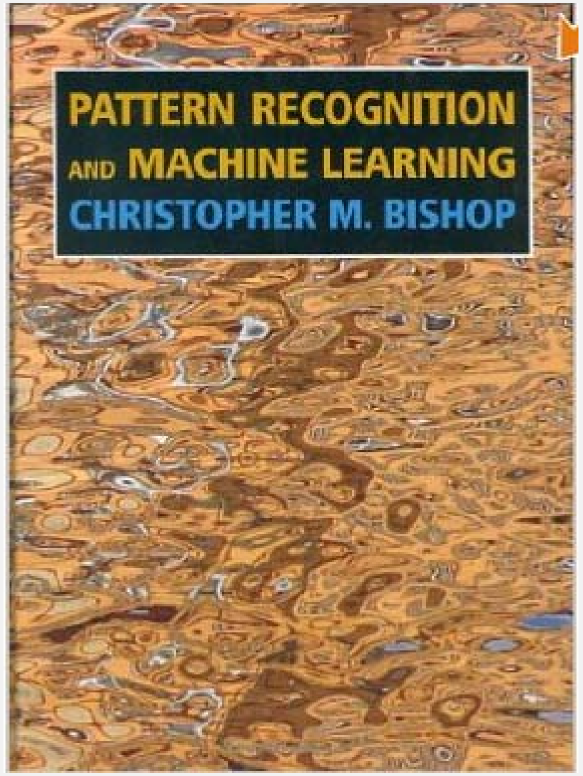
- Every CAS, as described in the beginning of this presentation, must have all these properties used:

ECAL 2013

FUNDAMENTALS OF COLLECTIVE ADAPTIVE SYSTEMS
FOCAS





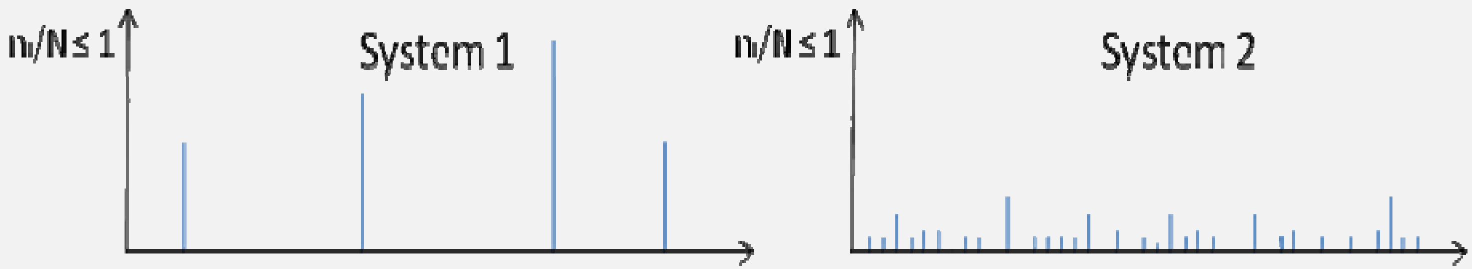


supplement:
by-hand examples

- $S = -\Sigma_i w_i \log w_i = -\Sigma_i (n_i/N)\log(n_i/N)$ $w_i \leq 1$ $\Sigma_i n_i = N$
- All we need is:
- $S = -\Sigma_i (n_i/N)\log(n_i/N)$
- We can take any log, here to the basis of 2, which does not change anything in principle:
- We take all i (persons, institutes, what's convenient), take the number of patents per i, calculate $-(n_i/N)\log(n_i/N)$ and sum over all i, that gives exactly one figure, the innovations entropy we want
- We have the extreme values:
  - lowest entropy → maximum order → all contents in one state → we know all outcome before → innovation is unlikely to happen
  - highest entropy → equal distribution over all states → maximum uncertainty of the outcome → maximal probability for innovation

supplement:
by-hand examples

- The figure below illustrates examples inside between the extreme values. This could be count of patents (to get from some national patent office) per person of a nation. Most persons don't have patents; they are simple empty slots just to some over (contributing 0) and to count for the total N. It is to mention, that entropy is an extensive thermodynamic variable, so doubling the systems means doubling the entropy, good for China, good for a big gene pool:



- Here we have:

  entropy S(System 2) = probability for innovation (System 2)

  ≥

  entropy S(System 1) = probability for innovation (System 1)

supplement:
by-hand examples

- Let's take $\log_2$ for below and in the following just for convenience for the examples:
  $S = -\Sigma_i(n_i/N)\log(n_i/N)$

- Let the slots be shown by | |
- and the n_i by |4|

- The maximum value is equal distribution, so in our simple example
- N = 8*2 = 16
  $S_{max}(|2|2|2|2|2|2|2|2|) = -8(n_i/N)*\log(n_i/N) = - 8*2/16*\log2/16 = -\log1/8 = \log8 = \log2^3 = 3$

- The minimum value is: just one guy has ideas (reminder: $\lim_{ni \to 0} n_i/N \ln n_i/N = 0$):
  $S_{min}(|0|0|0|16|0|0|0|0|) = -16/16*\log16/16 = 0$
  Significiant, that is the same for all variants of just one guy has ideas:
  $S_{also\ min}(|0|0|0|7|0|0|0|0|) = -7/7*\log7/7 = 0$
  That is an absolute necessary property:
- That is Nernst's theorem = 3rd law of thermodynamics

supplement:
by-hand examples

- Entropy is an extensive quantity, which means for example, doubling the system means doubling the value
- That is also an absolute necessary property!
- (An intensive quantity for example is the temperature. When doubling the system, the temperature stays the same)

- So from above:
  H(|2|2|2|2|2|2|2|2|) = 3
  H(|2|2|2|2|2|2|2|2|2|2|2|2|2|2|2|2|) = 6
  H(|2|2|2|2|2|2|2|2|2|2|2|2|2|2|2|2|) = 2 * H(|2|2|2|2|2|2|2|2|)

- Now let's take some arbitrary easy calculable example:
  $H_{between}$(|0|1|0|2|8|4|0|1|) = 1 + 7/8
  So we have:
  H_{max}(|2|2|2|2|2|2|2|2|) ≥ $H_{between}$(|0|1|0|2|8|4|0|1|) ≥ H_{min}(|0|0|0|16|0|0|0|0|)
  $Innavation_{max}$ ≥ $Innovation_{between}$ ≥ $Innovation_{min}$
- That does mean also, that one can reach the same entropy say innovation in two ways:
- By pure seize (bigger countries) or by a better distribution (smaller high-tech countries):
  H(|0|1|0|2|8|4|0|1|) = 15/8 = H(|1|2|8|4|1|)
- So pure size matters, the answer for the smaller ones is: Enhance entropy!