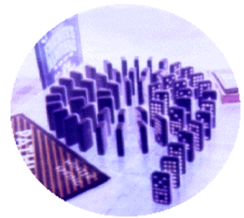
Thermal roots of correlation-based complexity



P. Fraundorf
UM-StL: Physics&Astronomy/Center for Molecular Electronics

Here I've collected some visualizations to help develop awareness of entropy as *lack of correlation* between system and environment. These may help deepen student insight into widening applications of information physics e.g. in molecular biology, computer science, even journalism. The *delocalized physical structure* of correlations provides further insight into the natural history of invention, and into emerging relationships between codes and excitations.

Hi. This talk is a review of helpful visualizations developed for discussing the thermal roots of correlation-based complexity. Predating my focus on using electron microscopes to examine VLSI silicon and extraterrestrial materials, this interest was inspired by the Tribus & McIrvine (1971) Scientific American article titled "Energy and Information", and many papers by E. T. Jaynes before I went to Washington University with an eye toward working with him. This is a review of connections that have been known in one form or another for a long time, but that may be unfamiliar to many so feel free to interrupt for clarification along the way.

Conclusions (non-graphic) in advance…

We review ways that Bayesian ``max-ent'' approaches, following Ed Jaynes and Myron Tribus, allow one to **integrate thermal physics and information theory points of view** into the quantitative study of complex systems.

Mutual information (general correlation between subsystems) is well known to be a special case of **net surprisal**, a free-energy analog that measures ``departures from expected'' in both thermal and information systems. **Information engines** manufacture these correlations by reversible thermalization of available work.

C**orrelations** between structures (e.g. a phenomenon and its explanation, or an organism and its niche) **within and across boundary types** (ranging from the edges of molecules to the gap between cultures) **are de-localized structures**. Development of such correlations is a big part of the natural history of invention, and categorization by boundary introduces us to **old concepts in a new guise**.

(adapted from arXiv:physics/9611022v3)

For those who can't stay, I'll first non-graphically outline where we're headed.

Gambling theory (MaxEnt) review

Multiplicity and it's log:	$S = k \ln \Omega$ in [nats, bits, or J/K] if $k \equiv [1, \frac{1}{\ln 2}, k_B]$
Uncertainty Slopes and Best-Guess Eqns of State:	X equilibrated $\Rightarrow S_{tot}$maxmized $\Rightarrow$ all $\frac{\partial S_i}{\partial X}$ equal $\Omega \propto U^{\frac{\nu N}{2}} \Rightarrow U = \frac{\nu N}{2} kT$ where $\frac{\partial S}{\partial U} \equiv \frac{1}{kT}$ $\Omega \propto V^N \Rightarrow PV = NkT = nRT$ where $\frac{\partial S}{\partial V} \equiv \frac{P}{kT}$
Multiplicity Exponents:	$\xi_U \equiv \frac{U}{kT} = U \frac{\partial S}{\partial U}$ and $C_U / k \equiv \frac{1}{k} \frac{\partial U}{\partial T} = T \frac{\partial S}{\partial T}$
Net Surprisal & dev. from expected (e.g. free energy over kT & mutual information):	$I_{net} \equiv -k \sum_{i=1}^{\Omega} p_i \ln(\frac{p_{oi}}{p_i}) \geq 0,$ $\frac{I_{net}}{k} = \frac{\nu}{2} N \Theta[\frac{T}{T_o}] + \sum_j N_j \Theta[\frac{N_{jo}}{N_j}]$ where $\Theta[x] \equiv x - 1 - \ln x \geq 0$

future links: story #1, #2, #3; puzzler #1, #2, #3; faq; read more about it

For each row above, note the relationship discussed, and its units.

(i) If we define surprisal as the log of probability's reciprocal for each accessible state, average surprisal (entropy, uncertainty) has information units (how many are familiar with this?);

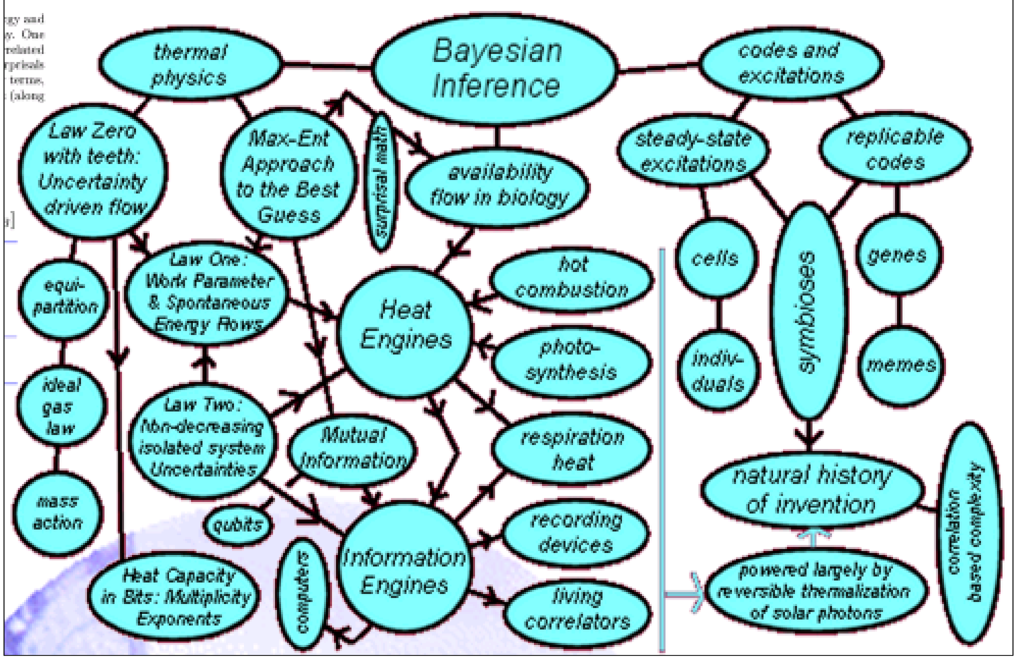
(ii) Max-ent best guesses (e.g. in micro-canonical, canonical, pressure and Gibbs ensembles) yield intensive Lagrange multipliers including reciprocal temperature, free expansion coefficient, etc. These are often derivatives of entropy wrt an extensive conserved quantity X that "equilibrates" as initial conditions fade. For example, 1/kT = dS/dE makes temperature an energy derivative that's not always proportional to energy in spite of occasional textbook allusions to the contrary (how many have seen this?);

(iii) Dimensionless integral and differential capacities (e.g. AJP 71, 1142) then have units of what? (Answer: bits per 2-fold increase in X or one of it's multipliers); and

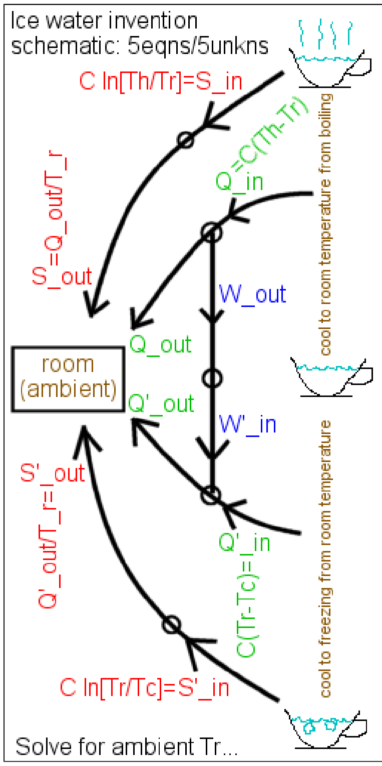
(iv) Net-surprisals (AKA relative or cross-entropy) in information units measure finite deviations from expected, and reduce near equilibrium to availabilities (free energy over kT) and in the case of correlated subsystems to mutual information (now fashionable in the study of evolving codes, nonlinear dynamics, and quantum computing). Bayesian inference also (cf. http://www.astro.cornell.edu/staff/loredo/bayes/) underlies many familiar distributions and estimators (e.g. Wiener phase PRL 64, 1031).

Note on multiplicity exponent partials: dS/dU, dU/dT and dS/dT may be no-work partials, and might also be defined with ensemble constraints e.g. if an appropriate enthalpy is allowed to substitute for U. This slide, by the way, is example of one kind of visual sound-byte (or light-meme) that might have a variety of auxiliary classroom uses, if linked to suitable supporting materials on the web.

Information physics hypermap



In addition to giving form to the laws of thermal physics, gambling theory shows how these laws have implications for available-work flow in biological systems. It provides ways to compress data, to track the evolution of chain letters and nucleic acid codes, and to provide context on the symbioses between replicable codes and steady-state excitations that underlie the natural history of invention, as well as distribution of the daily news. Rather than learn about the subject in specialized form in many different fields, a description informed to disparate applications at the outset might be worth exploration.



A UIUC intro-physics homework problem...

...solvable with three approaches:

- p-chem: Carnot engine and heat pump
- thermal physics: heat & entropy flows
- information physics: equal net-suprisals

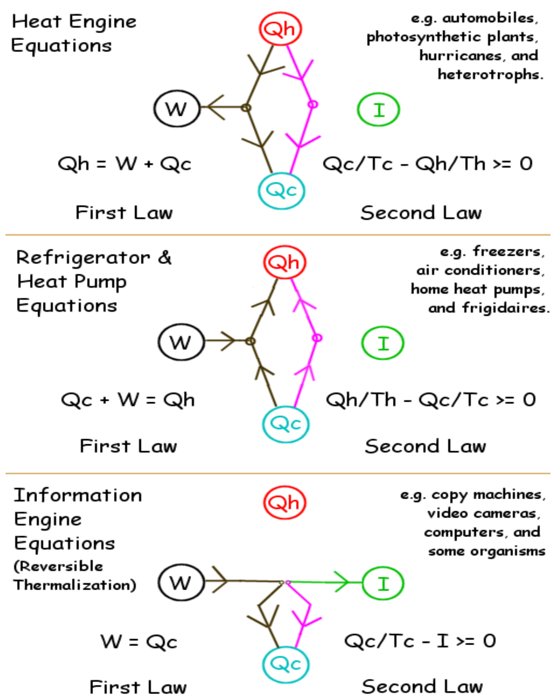
$$C^*\Xi[T_h/T_r]=C^*\Xi[T_c/T_r]$$

where...

$$\Xi[x]\equiv x-1-\ln[x]$$

A practical application of net surprisals involves a problem that my daughter was assigned in a physics course when she was a student here at U of I. Imagine a device that takes in a cup of boiling water, and outputs a cup of ice water. What is the maximum temperature at which such a device can operate without requiring an external source of available work? (Any suggestions here on what the answer is?)

I'm told that chemists might tackle this problem using Carnot engine and heat pump efficiencies, while physics students might use heat and entropy flows in introductory physics. However, by far the simplest approach is just to assume that an unpowered device can't increase the net surprisal of the water cup relative to ambient. Equating net surprisals (with respect to ambient) of boiling and ice temperature water, as shown above, then yields a simple equation which shows that such a device is possible even if the ambient temperature is well above 100F. This is one of many opportunities for future inventors that such calculations can uncover.

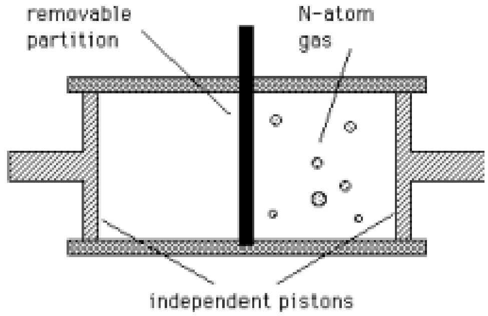


Graphical review of everyday engines

Heat engines and heat pumps are familiar from many introductory courses, but information engines that create correlations between subsystems by thermalizing available work are less familiar. They are, however, at least as important in our day-to-day lives.

The previous slide's approach to visualizing heat and entropy equations also works for heat engines and heat pumps. Information engines are less well known (primarily because thermodynamics traditionally avoids subsystem correlations). However, they are at least as ubiquitous.

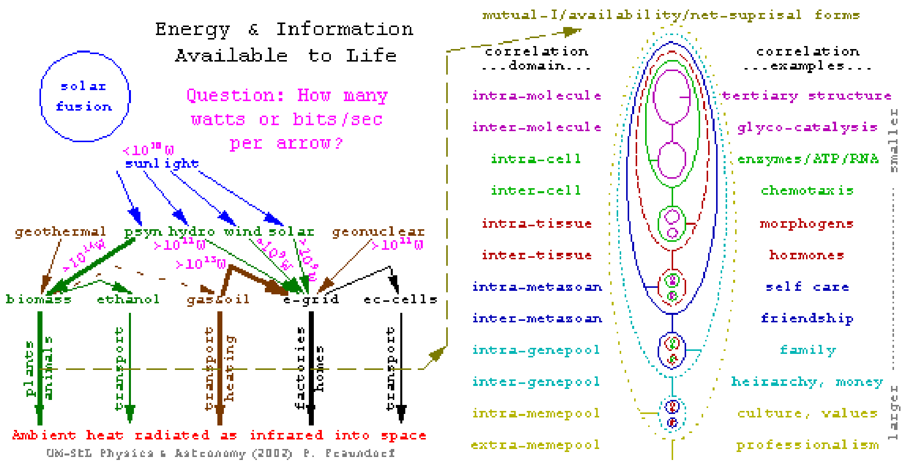
Natural units for temperature & entropy
facilitate the application of
thermal physics to information engines…



The price of correlating two subsystems (e.g. a code and and excitation) is for Szilard's vacuum-pump memory and in general: Work/I = $kT_{ambient}$ ln(2) [e.g. in joules/bit].

Above find a simple physical example of an information engine that can reversibly thermalize work to create a correlation between two subsystems. In the one-atom case, it can approach the thermodynamic limit for correlation-creation at a given ambient temperature. Copy machines and video cameras and even astronomers drawing maps of the night sky are other examples of such engines, which are also quite topical in modern day studies of code replication (cf. Tom Schneider at NIH), of code origins (cf. the June 2003 article by Bennett and others in Scientific American on chain letters), and of quantum computing (cf. articles in the past decade on mutual information and its applications by Seth Lloyd).

Life’s power stream / correlation-based complexity



Refs: cf. http://www.umsl.edu/~fraundor/hcapbits.html

The left hand side of this plot tracks the flow of available work through life on earth e.g. from incoming sunlight to its exit from our planet as thermal radiation. The right hand side offers an inventory by level of correlations within and across half a dozen sub-system boundary types, ranging from molecule surfaces to the boundary between cultures. Thus thermal physics in natural units provides clues to the role of reversible thermalization in the natural history of invention. Since respect for these correlations will increase in importance as social distance across the planet continues to shrink, information physics injects timely new life into the study of evolving correlation-based complexity.

Temporally-stacked layers of correlation-based complexity,
and the natural history of invention...

new level drivers	*boundaries*	*emerging correlations*
interstellar cloud collapse	radial temperature variation	spin up & stellar ignition
orbital accretion of dust & gas	radial pressure variation	planetary differentiation & geocycles
geothermal & solar gradients	compositional variation	biomolecular cycles
biological cells	bilayer membranes & cell walls	chemical communication, microbial symbioses & differentiation
biofilms & live tissues	organ surfaces	skeletal, respiratory, digestive, & nervous systems
metazoans	individual skins	pair bonds & redirected aggression
reproductive bargains, family	gene pool boundaries	social heirarchies & politics, ritualized available work
cultures & belief systems	meme pool boundaries	sciences & diversity protocols

The table above sketches the sequential development of some boundaries that one might use to categorize correlations. On the left you'll find physical processes that may be pre-requisite to the formation of these boundaries, and on the right you'll find physical processes that emerge after the boundaries are established. One might think of this sequential emergence of processes (and concepts associated with them) as a natural history of invention.

Note that each of these boundaries involves the distribution in space of physically-correlated structures. Consider the physical geometry of a gene-pool boundary. Yikes! Those geometers among us, who have been "spoiled" by getting to think about crystalline materials for example, are left in the dust by the geometric complexity of this beast. Thus our understanding about the detailed nature of code-pool boundaries in the last two entries, and the geometry of their distribution, is humbling. For visualization purposes, we cartoonify these boundaries symbolically later in the talk. Nonetheless, those symbols refer to quite physical molecules and memories correlated with other elements of their environment.

An *I*-physics, or correlation-based complexity, **boundary map**...

...to aid reference to processes that point in and out from molecule, cell, tissue, organism, gene-pool and meme-pool boundaries

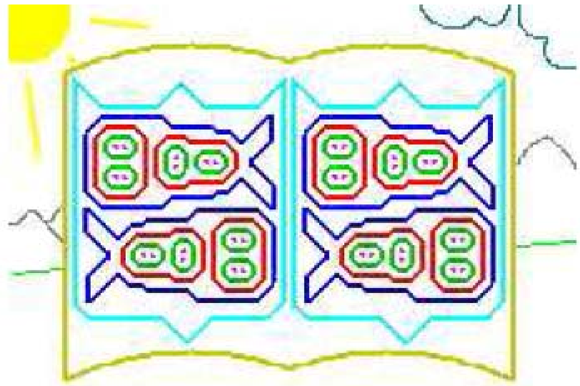
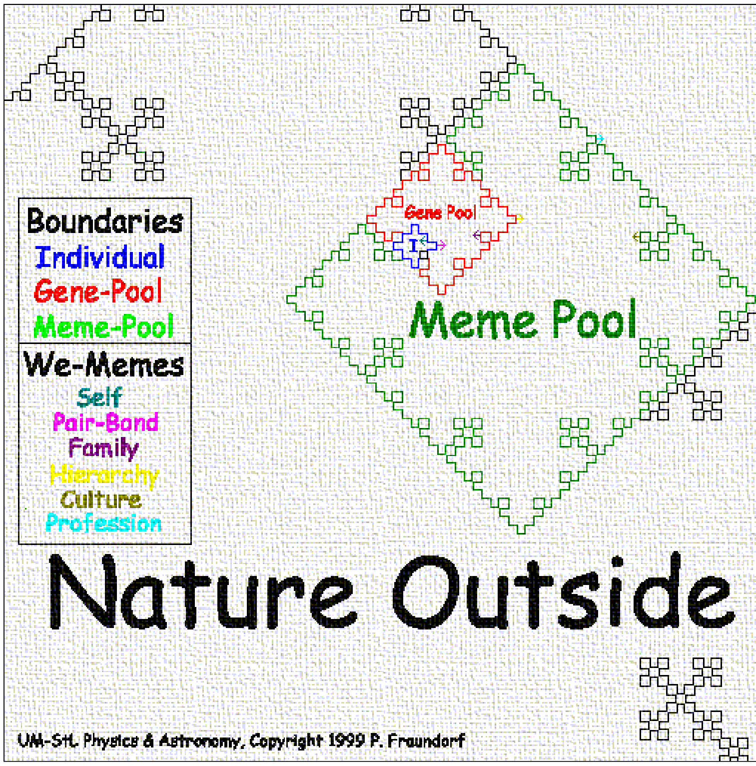


FIG. 6: An expanded schematic of some boundaries alone, drawn from Fig 5. This suggests for example that the idea sets for interaction between cultures (i.e. when one has to take into account more than one book in the figure, or meme-pool) may be quite different than those for interactions between heirarchies in a given culture. Guidelines for professionalism in the workplace, and political correctness, as irrelevant as they may seem in the context of one culture, might be evidence of behavior correlations developing along those lines.

A cartoon of some of the higher-level boundaries alone is I think quite mnemonic. In the figure above, the book outline represents the boundaries of a culture or meme-pool, the coat of arms outline represents the boundaries of one family or gene-pool, the fish outline represents the boundaries of an individual metazoan, the red outline the boundary of an organ subsystem, the green outline a cell membrane, and the magenta dots the edges of a molecule.

For example, in my own research (a niche of responsibility focusing outward from my cultural roots) I spend a lot of time exploring inter-molecule relationships. Like most humans, of course, I might also play a positive intra-cultural role e.g. in church activites and/or performing arts, and serve in a multi-family social hierarchy here and there e.g. at work and in the hood. A few intra-family responsibilities may pop up too, not to mention pair-bonds, etc. With others similarly engaged, organizational processes directed inward and outward from each of these boundaries, along with ideas designed to characterize and support them, might be expected to develop a life of their own. Indeed, it would seem that they have...



Ideas (replicable memetic codes) that help maintain these correlations also break down nicely in terms of their relationship to these boundaries.

For example individual, gene-pool, and meme-pool boundaries nicely classify the concept-sets used to describe and/or help maintain subsystem correlations with inward and outward focus from these three boundaries, in many different settings. These concepts of self-organization (or we-memes) fall into familiar categories. To wit: Ideas focusing inward and outward from organism boundaries include respectively, for example, self-maintenance and friendship, those focusing inward and outward from gene-pool boundaries include family and hierarchy, and those focusing inward and outward from meme-pool boundaries include culture and professionalism.

Niches of responsibility for such correlations have also evolved names. Working our way down from outward-looking meme-pool niches, for example, one might think of student and scholar, parishoner and priest, citizen and leader, sibling and parent, colleague and mentor, metazoan and medicine man. How about an evolved name for the correlation between the content of a code, and an aspect of the environment to which it pertains. Try: Is it true? The information physics connection here might thus inspire one to explore the possibility that respect for sub-system correlations on all of these levels, and the ability of individuals to participate therein, may help make the most of the limited flow of net-surprisal available to life on earth.

Living codes

This view of correlation-based complexity can facilitate dialogs ``above the dialog'' that recognize replicable codes (e.g. memes & genes) as well as steady-state excitations (e.g. organisms & machines) as **complementary** agents whose ``interest'' (e.g. survival or replication) factors into the way sub-system correlations develop.

You'll note that the public media focus pretty much exclusively on the perspective of organisms. Dialogs which also consider the perspective of replicable codes might in this sense be considered "meta-dialogs" of a sort that could be very helpful in preventing ideas (in effect busy serving their own replication today with world wide web assistance) from giving organisms short shrift. In other words, an ordinary dialog is a replicable code that does not explicitly address the way it does (or does not) serve its own interest. The meta-dialogs above are threads of discussion that, like humans when appropriate, explicitly discuss both code and organism perspectives including potential conflicts of interest associated with their own replication. A conflict of ``interest'' might arise, for example, when a story on the web inadvertantly exploits some reaction among people that result in wider and wider replication of the story itself.

Complex-system informatics

Courses for **code-based science** & **information industry** students that takes advantage of perspectives allowing one to paint with a unified-brush practical tools (like *mutual information*) for recognizing and measuring correlations between observations, and thus for example to:

- track evolutionary divergences in molecular and memetic codes,
- compress data/images and error-correct their transmission,
- understand processes converting ordered to thermal energy,
- find hidden attractors in non-linear dynamical systems,
- recognize complementarities crucial to the "big picture",
- inventory the standing crop of correlations in complex systems,
- monitor and maximize the content-bandwidth of media channels,
- optimize one's bets at the casino or in life, and
- predict physical limitations on quantum computing.

The importance of Bayesian inference in understanding complex systems, and the ``concept-backbone'' described here, suggest that a cross-disciplinary course on these connections particularly for students in the code-based sciences (e.g. molecular biology and computer science) and information industries (e.g. journalism) might benefit from an emergent content course in this area.

Recap (slides at http://www.umsl.edu/~fraundor)

• The second law extends beautifully to **correlations between subsystems** through a concept of net-surprisal first called to my attention in a 1971 *Sci Am* article by Tribus & McIrvine.
• **Net-surprisal** has useful if not well-known applications in thermal physics, e.g. to the ice-water invention problem.
• The *Szilard vacuum-pump memory* is a simple **information engine** that illustrates the thermodynamic cost of correlations between a physical system and a code keeping track of its state.
• **Life's power stream** channels thermodynamic availability into correlations directed inward and outward from *molecule*, *cell*, *organ*, *individual*, *gene-pool* and *meme-pool* boundaries.
• Energy, information, *steady-state excitations* and *replicable codes* repeatedly theme-up in the **natural history of invention**.
• Ideas that track correlations inward and outward with respect to these boundaries (i.e. **we-memes**) map to traditional categories!

2005 May Talk at UIUC: *The End*

Adventure: ....*The Middle*.........