Chapter 1

Net surprisals ala Tribus: correlations from reversible thermalization

P. Fraundorf
U. Missouri-StL Physics & Astronomy/CME
pfraundorf@umsl.edu

The Bayesian vision of net surprisals underlying the connection between energy and information, put forward by Myron Tribus four decades ago, offers a robust measure of correlation with "second law teeth". A special case is mutual information, whose widespread application to the study of correlated codes, quantum computing, and nonlinear dynamics is grounded in thermodynamic consideration of correlated subsystems. Surprisal physics can help students quantify finite departures from the ambient, and explain information engines which reversibly thermalize available work to literally power the natural history of invention. It might prove useful for tracking heirarchical correlations in complex systems, as well as code-excitation complementarity.

1.1 Introduction

"*Scientia potestas est* — 'Knowledge is power' — the Romans said, and 20 centuries later science has given an old phrase new meaning." Thus began Tribus and McIrvine's *Scientific American* article [Tribus and McIrvine(1971)] on *Energy and Information*. The article itself was a topical summary, not an exposition of original facts, but its composition drew from deep involvement in wider applications of Bayesian inference (cf. [Tribus(1961), Tribus(1969)]). It also drew from intuition concordant with future developments. We make this case with a brief look at some mathematics of Bayesian inference, based on notes provided by E. T. Jaynes, followed by description of a few possibly useful connections not yet been described in the literature. One objective is to show that a few decades after the first papers on Bayesian inference by Myron Tribus, science may be giving them new meaning as well.

1.2 Net surprisal's MaxEnt roots

A bit of surprisal is what you feel, after tossing one coin a few times for practice and then finding that your prediction of heads on the next toss is correct. A byte of surprisal (8 bits) is what you feel, after tossing eight coins a few times for practice and then predicting correctly in advance the resting state of all eight coins on the next toss. The more familiar quantity "probability" is simply $1/2^{\#bits}$, so surprisal goes (logarithmically) from 0 to ∞ as probability goes from 1 down to 0.

If each possible state of a system has equal surprisal, conditions which obtain for many states (e.g. 4 heads and 4 tails, instead of 8 heads) are more likely to happen by chance. This gives rise to the classical maximum entropy "best-guess" machine, which says that the most likely condition is that which maximizes average surprisal within constraints provided by all else that is known [Jaynes(1957a), Jaynes(1957b)]. The formalism itself is quite remarkable, because it allows one to develop familiar expressions for the first and second laws of thermodynamics from the mathematics of inference alone, before any physics (e.g. conservation of energy) is brought into the picture at all.

The "MaxEnt strategy" thus follows Tribus' prescription for Bayesian inference in general, which treats surprisal and probability assignments as"numerical encodings of states of knowledge" [Tribus(1969)] rather than as e.g. intrinsic properties of a particular subsystem. As we'll see, the connection between these quantities and "statements about correlation", e.g. between a prediction and a future behavior, is standing the test of time.

The prescient observation that we focus on in this paper involves remarks in the *Scientific American* article [Tribus and McIrvine(1971)] about the importance of differences between two entropies (referred to here as "net surprisals"), and their connection to free energy in units of kT, to Gibbsian availabilities, and to measures of finite departure from equilibrium. These connections also follow in detail from the maximum entropy prescription, cf. [Fraundorf(2003a)]. For purposes of this note we need only the mathematical definition of net surprisal or cross-entropy:

$$I_{net} \equiv -k \sum_{i=1}^{\Omega} p_i \ln(\frac{p_{oi}}{p_i}) \geq 0 \tag{1.1}$$

and the MaxEnt result that net surprisal under the exchange of conserved quantities is the entropy (or "average surprisal") increase of *system plus environment* on equilibration to ambient. Note that this prescription is designed for characterizing *finite* departures from ambient, and under broad conditions establishes "second law" relevance for the measure as well. Near ambient conditions, changes in availability (free energy in units of kT) also represent changes in net surprisal.

1.2.1 Thermal physics' ice water invention

Before we move on, it's worthwhile mentioning that net surprisals may be underutilized even in introductory thermodynamics. For example, the equations discussed in that *Scientific American* article simplify in the "ice water invention" problem, which asks how cold a room must be for an otherwise unpowered device to take boiling water in at the top, returning it as ice water with some ice therein at the bottom. Using $I_{net} \simeq C_v \Theta[T/T_{room}]$ where $\Theta[x] \equiv x - 1 - ln[x]$ and C_v is specific heat, it is easy to show that the second law requires only a room whose temperature is cooler than triple digits Fahrenheit, since boiling water has more than enough net surprisal to generate the needed chill.

1.3 Mutual information as a special case

One moves beyond the world of classical thermodynamics, with it's "extensive entropy", by including in the MaxEnt calculations information on the correlation between sub-systems e.g. knowledge that two data strings are identical even if we have no idea what they contain. Interestingly enough, this "mutual information" can written e.g. for two subsystems I and J as

$$M[I,J] = -k \sum_i \sum_j p[ij] \ln \frac{p[i]p[j]}{p[ij]} \geq 0 \tag{1.2}$$

i.e. as the net surprisal that follows upon learning that the two systems are not independent. Thus mutual information may (under some conditions at least) be subject to the same second law considerations that net surprisals obey more generally.

1.3.1 Information engine efficiencies

Returning briefly to introductory thermodynamics, an introduction to natural as well as historical units for thermal physics quantities [Fraundorf(2003b)] would allow teachers to introduce the concept of information engine along with the concept of heat engine. Information engines thermalize available work to create correlations between subsystems in their environment. The $\mathrm{Work}_{\mathrm{IN}}/\mathrm{kT}$ upper limit on the efficiency of such engines (whether they be digital cameras, copy machines, molecular motors, or an astronomer making notes on the night sky) is probably as fundamental, if not yet as useful, as the Carnot limit on the efficiency of heat engines.

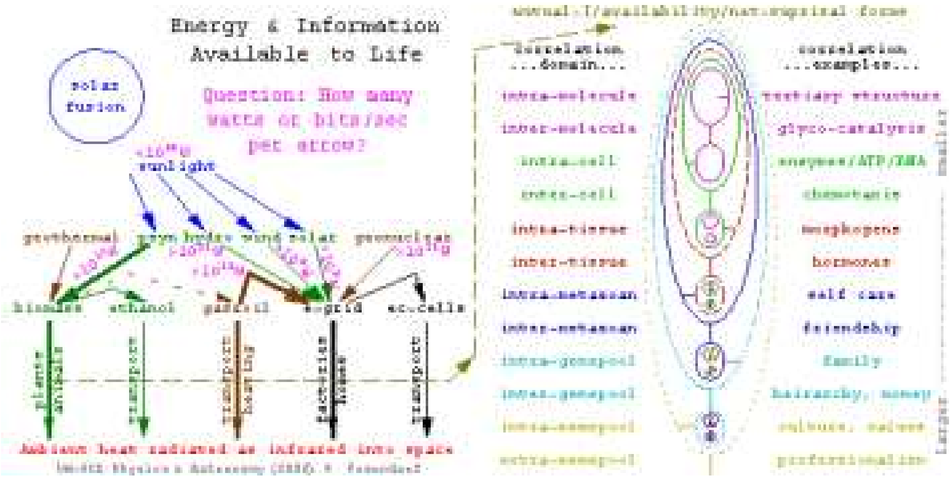

Figure 1.1: Life's Energy Flow (**Left**): The top half represents primary processes involved with energy flow; the bottom half physical repositories for life's energies, and paths for conversion of energy from one form to another. Life's Stores of Availability (**Right**): Horizontal bars represent inward-looking correlations; vertical bars outward-looking correlations. This breakdown categorizes by domain the types of correlations that exist, and types of memetic replicators that help to maintain them.

1.3.2 Energy and information available to life

The left half of Figure 1.1 illustrates the flow of energy available to life on earth. Much of what begins as radiative heat from the sun, is converted to chemical potential energy (e.g. heat engine style) in the form of plant biomass [Odum(1971)]. The flow rate through biomass is probably worth closer consideration [Vitousek et al.(1997)]. Many introductory physics texts and even the world almanac, for example, ignore its size entirely. Thus student projects on *the size* of these flows, at various times and places, might provide useful consumer education.

Some of the "ordered energy" outputs from the heat engines described in Figure 1.1 are eventually thermalized (e.g. in forest fires or the burning of fossil fuels), but not all of it is irreversibly thermalized. In other words, some of the free energy made available by plants is converted to non-energy related correlations between organism and environment, and some is converted to to internal correlations within living things.

Organism/environment correlations include, for example, cell membranes that separate the contents of one-celled organisms from the fluids surrounding. They differ, depending on the nature of the ambient to which a given organism is adapted. Similarly, the woody trunks of trees don't merely store chemical energy for later combustion, but instead point in a direction which allows subsequent leaf growth to have better access to the light of the sun.

Correlations internal to organisms include catalysts (often amino acid enzymes) which guide the spending of the cell's energy coin [Hoagland and Dodson(1998)] (adenosine triphospate molecules) not only toward nourishment and other external goals, but toward it's use in the process of cell replication. The enzymes themselves are typically constructed from amino acid sequences which fold in solution into secondary and delicate tertiary structures which are crucial to catalyst structure and function.

Information on these correlations within catalyst molecules, resulting in part also in correlations between organisms and their environment, apparently proved so important that a *digital* means (nucleic acid codes) to store mutual information on these correlations, still in widespread use today, was developed several billion years ago [Margulis and Sagan(1986), Ward and Brownlee(2000)]. Note that the word digital here refers to ways to store mutual information in which bit-wise fidelity of the replication process can be checked after the fact. This is distinct from *analog* forms of recording, like the storage of images on film, where accuracy on the microscopic scale is lost statistically in the grain structure of the film, as one moves to increasingly smaller size scales.

This ancient invention of digital recording more or less formalized a now long-standing symbiosis between *steady-state excitations* (in particular organisms which operate in-part by reversibly thermalizing an inward-flowing stream of energy in the form of available work) and *replicable codes*. This excitation-code symbiosis, of course, involves mutual information managed (stored, replicated, and applied to enzyme manufacture) by biological cells [Hoagland and Dodson(1998)]. Now memetic replicators [Dawkins(1989), Blackmore(2000)], i.e. ideas which began as sharable patterns stored in the neural nets of animals [Dennett(1992), van Schaik et al.(2003)], are in the process of going digital [McLuhan(1962), Harnad(1991)], thus adding a second level to life's symbiosis with replicators. The unconscious struggle for hegemony over organisms, between these two replicator families, might in a way be seen as a battle between "sword" and "pen" in which (strangely enough) organisms are the spoils of war. At the very least it seems likely that under some conditions the interests of organisms, and the interests of codes, are not one and the same.

A natural way to "illustrate and inventory" the standing crop of correlations associated with life, while recognizing boundaries between replicator-pools as well as simpler physical boundaries (like cell walls and individual spaces), is illustrated in the right half of Figure 1.1.

1.3.3 Invention's natural history

Histories of emergent phenomena, like Marshall McLuhan's "Gutenberg Galaxy" [McLuhan(1962)], Konrad Lorenz' "On Aggression" [Lorenz(1966)], Margulis & Sagan's "Microcosmos" [Margulis and Sagan(1986)], Jared Diamond's "Guns Germs and Steel" [Diamond(1997)], David Attenborough's Special on Birds [Attenborough(1997)], and Ward & Brownlee's "Rare Earth" [Ward and Brownlee(2000)] (in broad strokes at least) simplify when outlined in terms of the two manifestations of generalized availability depicted in Figure 1.1: (i) "ordered" or free energy, and (ii) mutual information, respectively. These two themes repeatedly intertwine in a non-repeating drama that involves partnership between replicable codes (which incidentally include the above two concepts), and what physicists might call steady-state excitations busily converting energy and information from one form to another.

The natural history of invention's "time-line of concept-relevance" [Fraundorf(2003a)] suggests that correlations written in nucleic or amino acid strings have been developing in symbiosis with microbes since the very early days of our planet. Moreover, sometime since the Cambrian bloom of metazoan body types, and particularly among humans in the past 10,000 years, similar correlations written in memetic codes have been undergoing active development. The latter were broadcast not via the sharing of molecules, but by transference between neural nets through metazoan senses via performance, speech, script, and more recently digital means.

The large number of thermodynamic and information-theoretic processes found in the natural history of invention raises a familiar question about codes: What gene is responsible for what features of an organism, or conversely what features of an organism does a given nucleic acid sequence "cause"? The same question can be asked about memetic codes. Has a given set of ideas been honed by experience with the world around us, via experience with worlds within some parochial boundary, or does it offer little by way of connection to the world at all? Accepting Tribus' maxim about "states of knowledge", in any rigorous sense such questions are about delocalized correlations between physical objects (e.g. between a code and it's phenotype, or some other part of the world around) rather than about the properties of a molecule or a set of words in isolation. Once the context is specified, e.g. the reference state used in equation (1.1), objective and even quantitative assessments of such correlations may be possible.

Qualitatively, for example, most might agree that the nucleic acid base triplets UAA, UAG, and UGA have evolved as elements of punctuation in the genetic code, there not to correlate with the outside world but to guide the process of transcription into protein [Hoagland and Dodson(1998)], much as the period at the end of this sentence guides the sentence's transcription into speech. Such punctuation codes are one kind of internal code, developed to guide code replication and their reduction to practice. Other codes have evolved by virtue of the correlations that they affect between an organism and the inanimate world around. Thus a chunk of genetic code might correlate with the thickness to length ratio of a plant's stem, whose optimum value may depend on wind velocities and topography in the world around. Similarly, a set of ideas for guiding the path of a ship at sea might survive depending on its usefulness in helping the sailors reach their destination, before they run out of supplies.

1.4 Discussion and Summary

Consideration of gene, plus organism, perspectives is now accepted as crucial to understanding evolution [Dawkins(1989)], just as the perspective of idea replication, plus that of individuals, is helpful in understanding culture [McLuhan(1962), Blackmore(2000)]. This complementarity may arise because excitation-focus can result in irreversible thermalization (i.e. an out-of-control burn) while fixation with specific codes can result in a code's failure to adapt.

As tools in the "science of the possible", the net surprisals discussed in *Scientific American* [Tribus and McIrvine(1971)] may help tantalize engineers into creating the ice water invention of introductory thermodynamics. Awareness of mutual information, a special case of net surprisal, is already crucial to our understanding of quantum computers [Gottesman and Lo(2000)], and the molecular machines that keep us going today [Schneider(2000)]. Net surprisals further suggest a physical framework for considering such disparate (and sometimes competing) issues as conservation of available work, maintenance of genetic diversity within and between species, and the future of cultural diversity in the presence of evolving constraints. By shifting focus to the role of reversible thermalization in the natural history of invention, this provides useful context and limits on widely diverse systems. Thus the ideas discussed by Myron Tribus decades ago may be shown to carry new meaning in the days ahead.

1.5 Acknowledgments

Thanks go to Myron Tribus and the late E. T. Jaynes for many interesting papers and discussions; to Tom Schneider at NIH for spirited questions and comments.

Bibliography

[Tribus and McIrvine(1971)] M. Tribus and E. C. McIrvine, Sci. Amer. **224**(3), 179 (1971).

[Tribus(1961)] M. Tribus, *Thermostatics and thermodynamics* (D. Van Nostrand Co., Princeton, 1961).

[Tribus(1969)] M. Tribus, *Rational descriptions, decisions, and designs* (Pergamon Press, New York, 1969).

[Jaynes(1957a)] E. T. Jaynes, Phys. Rev. **106**, 620 (1957a).

[Jaynes(1957b)] E. T. Jaynes, Phys. Rev. **108**, 171 (1957b).

[Fraundorf(2003a)] P. Fraundorf, *Information physics: From energy to codes* (2003a), `arXiv:physics/9611022`.

[Fraundorf(2003b)] P. Fraundorf, Am. J. Phys. **71**(11), 1142 (2003b), `arXiv:cond-mat/9711074`.

[Odum(1971)] E. P. Odum, *Fundamentals of Ecology* (W. B. Saunders, Philadelphia, 1971), 3rd ed.

[Vitousek et al.(1997)] P. M. Vitousek, H. A. Mooney, J. Lubchenco, and J. M. Melillo, Science **277**, 494 (1997).

[Hoagland and Dodson(1998)] M. Hoagland and B. Dodson, *The way life works: A science lover's illustrated guide to how life grows, develops, reproduces, and gets along* (Three Rivers Press, New York, 1998).

[Margulis and Sagan(1986)] L. Margulis and D. Sagan, *Microcosmos: Four billion years of evolution from our microbial ancestors* (Simon and Schuster, New York, 1986).

[Ward and Brownlee(2000)] P. D. Ward and D. Brownlee, *Rare earth: Why complex life is uncommon in the universe* (Copernicus, New York, 2000).

[Dawkins(1989)] R. Dawkins, *The Selfish Gene* (Oxford University Press, New York, 1989), 2nd ed.

[Blackmore(2000)] S. Blackmore, *The Meme Machine* (Oxford University Press, 2000).

[Dennett(1992)] D. C. Dennett, *Consciousness Explained* (Little Brown and Co., 1992).

[van Schaik et al.(2003)] C. P. van Schaik, M. Ancrenaz, G. Borgen, B. Galdikas, C. D. Knott, I. Singleton, A. Suzuki, S. S. Utami, and M. Merrill, Science **299**, 102 (2003).

[McLuhan(1962)] M. McLuhan, *The Gutenberg galaxy: The making of typographic man* (publisher, address, 1962).

[Harnad(1991)] S. Harnad, Public-access Computer Systems Review **2**(1), 39 (1991).

[Lorenz(1966)] K. Lorenz, *On Aggression* (Harcourt, New York, 1966).

[Diamond(1997)] J. Diamond, *Guns, germs and steel: The fates of human societies* (Random House, New York, 1997).

[Attenborough(1997)] D. Attenborough, *Attenborough in paradise* (1997), a PBS special on song, dance, theatre, and bauer among birds on islands of New Guinea, photographed by Richard Kirby and Mike Potts, and produced by the BBC Natural History Unit.

[Gottesman and Lo(2000)] D. Gottesman and H.-K. Lo, Physics Today **53**(11), 22 (2000), `quant-ph/0111100`.

[Schneider(2000)] T. D. Schneider, Nucleic Acids Res **28**(14), 2794 (2000), URL `http://www-lmmb.ncifcrf.gov/ toms/`.