ISSN 0980-1472

Revue Internationale de

systēmique

Vol. 2, N° 1, 1988

afcet

Dunod

AFSCET

Revue Internationale de
systēmique

Revue Internationale de Sytémique

volume 02, numéro 1, pages 29 - 46, 1988

Towards a relativistic information theory for pattern and form

Guy Jumarie

Numérisation Afscet, janvier 2016.

BURNET F.M., The clonal selection theory of acquired immunity. Vanderbilt University Press, Nashville, 1958.

JERNE N.K., Towards a network theory of the immune system. *Annales d'immunologie*, Institut Pasteur, 125C, 1974, 373-389.

LYMAN F.L. Ed., Phenylketonuria, Thomas, Springfield, 1963.

MARCEL A., : Conscious and unconscious perception : experiments on visual masking and word recognition, *Cognitive Psychology*, 15, 1983, 197-237.

NEEDHAM J. : Biochemistry and Morphogenesis, Cambridge University Press, Cambridge, 1942.

SUTHERLAND E.W., PYE I. et BUTCHER R.W. : The action of epinephrine and the role of the adenyl cyclase system in hormone action, *Recent Progr. Horm., 21*, 1965, 623-658.

VARELA F. : *Principles of biological autonomy*, North Holland, 1979.

VAZ N. et VARELA F. : Self and non-sense : an organism-centered approach to immunology, *Medical Hypothesis 4*, 1978, 231-267.

WINOGRAD T. : What does it mean to understand language, in D. Norman (éd.) : *Perspectives on Cognitive Science*, Ablex : Norwood N.J., 1981.

REVUE INTERNATIONALE DE SYSTEMIQUE
Vol. 2, N° 1, 1988, pp.29 à 46

TOWARDS A RELATIVISTIC INFORMATION THEORY FOR PATTERN AND FORM [1]

Guy JUMARIE

Université du Québec à Montréal [2]

Abstract
This paper outlines a theory of relative information (that is to say with syntax and semantics) for pattern and form. Basically it combines two results, i.e. the randomization technique which previously provided us with a unified approach to discrete entropy and continuous entropy, and the so-called model of observation with informational invariance ; both of them are applied to a classical equation of Shannon information theory. One so derives measures of relativistic uncertainty for stochastic continuous processes and deterministic mappings, which are fully consistent with practical experiments on the one hand, and with the concept of fractal dimension on the other hadn. Prospects are outlined.

Résumé
Cet article définit les éléments d'une théorie relativiste de l'information (c'est-à-dire avec syntaxe et sémantique) contenue dans une forme. Essentiellement, il combine deux techniques, à savoir la méthode de randomisation qui nous permit d'obtenir une approche unifiée des entropies discrètes et continues, et notre modèle d'observation avec invariance informationnelle ; elles sont simultanément appliquées à une équation bien connue dans le formalisme de Shannon. On obtient ainsi des mesures de l'incertitude relative contenue dans une forme stochastique ou déterministe, qui sont, par ailleurs, pleinement en accord des résultats expérimentaux obtenus par d'autres chercheurs d'une part, et avec la notion de dimension fractale d'autre part. On esquisse ainsi un cadre qui pourrait permettre une étude générale des formes.

1. Recherche subventionnée par le Conseil National de la Recherche du Canada.

2. Département de Mathématiques et d'Informatique, B.P. 8888, St A, Montréal, QUE, H3C 3P8, Canada.

Revue internationale de systémique. Vol. 2/88/01/29-46/18/$ 3.80/© Gauthier-Villars

1. Introduction

Nyquist [17] is probably the first author who considered the problem of efficiency of transmission ; later Hartley [6] proposed the first known measure of information involved by a signal ; and in 1948 Shannon [18] stated his «Mathematical Theory of Communication» in its practical final form. Significant basic contributions, in these technical concerns, have been derived later by Gallager [5].

Very earlier Ashby [2] and Brillouin [3] pointed out that Shannon theory as so defined needed to be revisited in order to be fruitfully applied to natural sciences and to cybernetics : namely, it should be modified to deal with syntax and semantics (remark that Shannon himself already claimed that his formulation intentionally did not refer to semiology), but at date there were no significant advance in this way. Recently, extensive researches have been carried out about the question, and one can mention the approach via the *possibility theory* [6] and our *relativistic or relative information theory* which introduces a quantitative approach to subjectivity in information.

Indeed in 1975 [8], in an approach to a general system theory [11], we initiated this reltive information modelling [9] which can now be considered as being in its final form [15]. But a question which remained open until now is the problem of defining the entropy of a form or of a pattern, the solution of which is a prerequisite to cybernetics. Recently, we derived some preliminary results related to this question [12] and our purpose herein is to carry on this study.

The paper is organized as follows. For the convenience of the reader, we briefly recall the essentail of relative information on the one hand, and of our new concept of *total or complete entropy,* and then we shall refine our concept of *entropy of stochastic trajectories*. Lastly we shall introduce the relative feature of information in this framework, and we shall outline a possible extension to the entropy of deterministic mapping. Amazingly we shall cross over the theory of fractal dimensions initiated by Mandelbrot [16].

2. A few prerequisites

We believe that a summarized background on some results we previously obtained is mandatory since they will be applied to the derivation of the entropy of form and pattern.

2.1. *Observation with Informational Invariance*

The framework. Assume that an observer R is observing two continuous random variables $X \in R$ and $X' \in R$; and let p(x,x') denote the probability density of (X, X'). We shall assume that the amount of uncertainty involved by the pair (X, X') is measured by the Shannon entropy H(X, X'),

$$H(X,X') := \int_{R^2} p(x,x') \ln p(x,x')\, dxdx' \qquad (2.1)$$

It is well known that incidentally, but incidentally only, H(X,X') defines too the amount of information involved by the informational source (X,X').

Observation with informational invariance. We shall assume that the main features of the observation process ($R \rightarrow (X,X')$) are suitably described by the following axioms.

(A1) Due to coupling effects between X ans X', R cannot measure the exact values of x and x' respectively, but instead, the observation process results in an observed pair (x_r, x'_r) defined by the linear transformation

$$x_r = ax + bx' \qquad a,b \in R \qquad (2.2.)$$

$$x'_r = cx + ex' \qquad c,e \in R \qquad (2.3.)$$

(A2) The observation of the pair (x,x') does not create nor destroy the potential information involved by (X,X') considered as a source of information.

It is a simple matter to show that there are only two transformations which satisfy (A1) and (A2), and they are respectively defined by the equations

$$x_r = x \cos\theta + y \sin\theta \ , \ \theta \in R \qquad (2.4.)$$

$$x'_r = -x \sin\theta + y \cos\theta \qquad (2.5.)$$

and

$$x_r = x \cosh\omega + y \sinh\omega \ , \ \omega \in R \qquad (2.6.)$$

$$x'_r = x \sinh\omega + y \cosh\omega \qquad (2.7.)$$

Minkowskian observation. Assume now that the additional following axiom is satisfied,

(A3) In the special case where $x \equiv x'$, that is to say when R observes one variable only, then the equations of the observation process should reduce to

$$x_r = Q\, x$$

where Q denotes a constant gain cœfficient.

Proposition 2.1. Assume that the observable (x,x') is observed in such a way that Axioms (A1, 2, 3) are satisfied. Then the corresponding observation process referred to as *Minkowskian observation* is defined by equations (2.6) and (2.7) which can be re-written in the form

$$x_r = p(u)(x + ux') \qquad (2.8.)$$

$$x'_r = p(u)(x' + ux) \quad \square \qquad (2.9.)$$

Scaling factor. Assume that, on a physical standpoint, the variables X and X' are expressed by means of the same measurement unit ; then, by introducing a scaling factor c such that X and cX' have the same physical dimension, we shall then have the well known equations

$$x_r = p(u)(x + ux') \qquad (2.10)$$

$$x'_r = p(u)\left(x' + \frac{u}{c^2}x\right) \qquad (2.11.)$$

For further details about the practical meaning of the axioms (A1, «,3) above, see for instance the references [9, 13, 15].

Relativistic information. When the observed variables are given measures of information, one so obtains the so-called relativistic information modelling.

2.2. *Background on Total Entropy*

It is well known that continuous entropy and discrete entropy (in Shannon sense) are different in their mathematical natures in the sense that the former is not the limiting form of the latter when the discretizing span tends to zero. Mathematicians customarily claim that it is a matter of absolute continuity of a measure with respect to another one, but in a physical framework such an explanation is not fully satisfactory at all, and it is rather obvious that it is merely the practical meaning of these entropies which are different. So, in order to exhibit this difference, we introduced a concept of *total – or complete entropy* which we shall herein bear in mind.

The framework. Consider a discrete random variable X which takes on the values $(x_1, x_2, \ldots, x_m)$ with the respective probabilities $p_1, p_2, \ldots, p_m$. With each x_i, we associate an interval length h_i as pictured in Fig. 1, for the special illustrative case m = 3.

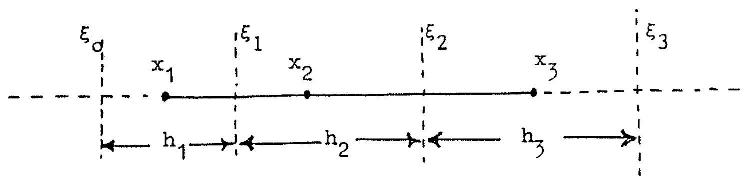

Fig. 1. Definition of Total Entropy

The min axioms. We shall assume that the total amount of uncertainty $H_e(X)$ involved by X should satisfy the following axioms.

(B1) $H_e(X)$ is a function $X[p_1, h_1), \ldots, (p_m, h_m)]$ the value of which is not modified by any permutation on the set $(p_1, h_1), (p_2, h_2), \ldots, (p_m, h_m)$.

(B2) X(.) is continuous w.r.t. $p_1, p_2, \ldots, p_m$ and $h_1, h_2, \ldots, h_m$.

(B3) X(.) is an increasing function of h_i for every i.

(B4) Let $\Phi(p_1, p_2, \ldots, p_m) = H(X)$ denote the Shannon entropy of X, then one has

$$x[(p_1, 1), \ldots, (p_m, 1)] = \Phi(p_1, p_2, \ldots p_m) \qquad (2.12.)$$

(B5) Let $(q_1, h'_1), (q_2, h'_2), \ldots, (q_n, h'_n)$ denote another random variable, then one has the equation

$$x[(p_1q_1, h_1h'_1), \ldots, (p_iq_j, h_ih'_j), \ldots, (p_mq_n, h_mh'_n)] =$$

$$= x[(p_1, h_1), \ldots, (p_m, h_m)] + x[(q_1, h'_1), \ldots, (q_n, h'_n)] \qquad (2.13.)$$

We then have the following result.

Proposition 2.2. A measure of uncertainty referred to as *total (or complete) entropy* which satisfies axioms (B1) to (B5) is defined by the expression

$$H_e(X) = H(X) + K \sum_{i=1}^{m} p_i \ln h_i \qquad (2.14.)$$

$$= -k \sum_{i=1}^{m} p_i \ln(p_i / h^{K/k}) \qquad (2.15.)$$

where K denotes a positive constant associated with the uncertainty involved by the intervals $(x_{i+1} - x_i)$. $\square$

Proposition 2.3. The continuous (Shannon) entropy is the limiting form of the discrete (Shannon) entropy when the discretizing span tends to zero. $\square$

For further details, see Ref [13].

Bibliographical Comments. In 1975 [8], by using physical arguments, we suggested a quite similar entropy referred to as «effective entropy» in an attempt to derive a modelling of negative information. Clearly, we defined

$$H_e(X) := H(X) + \sum_i p_i h_i(x_i)$$

where $h_i(x_i)$ is a uncertainty associated with the definition of the state x_i itself. Later Aczel and Daroczy [1], via the axiomatization technique of mathematicians, re-discovered exactly this same entropy that they re-named «inset entropy» [1].

3. On the entropy of form and pattern

3.1. *Human Feature Extractor*

The following result has been put in evidence by a large number of practical experiments and can then be taken for granted as being soundly supported.

When a human observer is examining a form, his cortex selects the salient features of this pattern, mainly the angles, and then uses them later to build-in a referential model. For instance, it has been proven that the nose of a face is of paramount importance in the identification of a portrait by a human observer. The basic conclusion which results from this remark is the following one.

Any recognition process of a form by a human cortex can be decomposed into two stages : (i) the first step in which the cortex characterizes the pattern by means of a *finite number of local features*, and (ii) the second step in which each local feature is individually analyzed one at a time.

The first stage involves a discrete entropy , the second stage deals with the amount of information contained in a continuous pattern, and we are going to comment on this point in the next sub-section.

3.2. *Classification of Continuous Observation Processes*

One may of course aim to define the entropy of a continuous pattern in an absolute way (it would then be a somewhat maximum possible uncertainty involved by this form) but it is much more realistic to consider that this uncertainty should depend upon how the form is observed ; and on doing so, we are more or less implicitly introducing the basic relative feature of information via the definition of the corresponding entropy. In this way, one can define three main types of observation for a stochastic scalar valued process.

White observation. In this process, we observe each point $(t,X(t))$ irrespective of the other ones, as if it were alone. Geometrically speaking, we intersect the trajectory F of $X(t)$ by a straightline parallel to the X-axis and we consider the position of the points so obtained. The uncertainty of the trajectory over a given finite horizon (t_o,t_f) is then the combination (in some sense to define) of the individual entropies for each $X(t)$.

Local Markovian observation. Assume that the trajectory is discretized in the finite sequence $X(t_o)$, $X(t_1)$, $X(t_2)$,...,$X(t_f)$; then the observer observes the corresponding trajectory in the form of the sequence of the pairs $(X(t_o),X(t_1))$, $(X(t_1),X(t_2),X(t_3))$,... The total uncertainty so involved by the trajectory to the observer is the combination in some sense to define of the various uncertainties involved by the different pairs $(X(t_i),X(t_{i+1}))$.

Total Markovian observation. In the model, each point $X(t_i)$ is observed respectively to the preceding one $X(t_{i+1})$. We then have the finite sequence $X(t_o)$, $(X(T_1)/X(t_o))$, $(X(t_2)/X(t_1))$,..., and the total amount of uncertainty so involved by the trajectory is the combination, in some sense to define, of the different uncertainties involved by the different conditional variables : $(X(t_i)/X(t_{i-1}))$.

1. L'«inset entropy» de Aczel et Daroczy admet, dans un cas particulier, une formulations identique à celle de l'«effective entropy» de Jumarie (Note du Comité de Rédaction).

4. Entropies of stochastic process

As we point out above, the uncertainty involved by a stochastic depneds upon how the latter is observed, and in the present section we give some expressions of this entropy under different modes of observation. For the sake of length we do not explain the derivation of these equations, transferring the reader to our main reference (Jumarie 1988) for further explanations.

Result 4.1. *White observation* Let $X(t) \epsilon\ R$ denote a continuous stochastic process with the probability density $p(x,t)$, and assume that the corresponding trajectory F is subject to a white observation by an observer R. Then, the uncertainty $H(F;O,T)$ so involved by F over the time range (O,T) is defined by the expression

$$H(F;0,T) := -\frac{1}{T}\int_o^T \int_R p(x,t)\ln p(x,t)dxdt + \ln T \qquad (4.1.)$$

Result 4.2. *Non white observation.* Assume that the stochastique dynamics $X(t) \epsilon\ R$ is observed by an observer in a non white observation process. We shall infer that this observation is characterized by a kernel $u(t) \geqslant 0$ in such a manner that the amount of uncertainty so involved by the trajectory F over the time range $(0,T)$ is measured by the expression.

$$H(F;0,T) := -\frac{1}{T}\int_o^T \int_R u(t)\,p(x,t)\ln p(x,t)dxdt + \ln T \qquad (4.2.)$$

In the special case of the Markovian process defined by the equation.

$$\dot{x}(t) = (f(x,t) + g(x,t)w(t) \tag{4.3.}$$

where w(t) is a white noise with zero mean and unit variance, one has

$$H(F;0,T) = \frac{1}{T}\int_o^T \int_R p(x,t) \ln \frac{p(x.t)}{|g(x,t)| (2\pi e)^{\frac{1}{2}}} dxdt \tag{4.4.}$$

Well obviously, this expression represents too the trajectory enropy of a Markovian process under local Markovian observation.

Result 4.3. The trajectory entropy under global Markovian observation of the Markovian process defined by Equation (4.3.) is

$$H(F;0,T) = H(X;0) + \frac{1}{T}\int_o^T \int p(x,t) \ln [\, |g(x,t)| (2\pi e)^{\frac{1}{2}} \,] dxdt + \ln T \tag{4.5.}$$

5. Relative entropy of stochastic trajectory

5.1. *On the Level of Subjectivity*

Basically, the subjectivity is introduced at the level of the observation of the trajectory by the observer. If this trajectory is observed in its entirety via the selection of a finite number of striking features, then the usual framework of relative information applies directly. In contrast, in a Markovian observation, this subjectivity should appear locally, and the overall subjectivity would then be the combination, in some sense to define, of these local subjectivities. In other words, it is the observation of the entropy H(X;t) which then involves subjectivity.

5.2. *Relative Trajectory Entropy*

Let $X(t) \in R$ and $X'(t) \in R$ denote two stochastic processes, and assume that their respective entropies H(X;t) and H(X';t) are observed in such a way that the axioms A1, A2 and A3 of section 2 are satisfied. As a result, this local observation provides a relative entropy in the form

$$H_r(X;t) = p[u(t)]\,[H(X;t) + u(t)H(X';t)\,] \tag{5.1.}$$

In short, we so associate with each variable X(t) another one X'(t)

which can be thought of as picturing the practical meaning of X(t) : in other words, *H(X';t) is the semantical entropy of X(t).* We then have the following result.

Proposition 5.1. Assume that the pair (H(X;t), H(X';t)) is observed following a Minkowskian observation to yield the local relative entropy expressed by equation (5.1.) ; then in the framework of the randomization with respect to time, the corresponding *relative trajectory entropy* H_r (F;0,T) is defined by the equation

$$H_r(F;0,T) = \frac{1}{T}\int_o^T \mu(t)\rho[u(t)]\,[H(X;t) + u(t)H(X';t)\,]\,dt + K \ln T \quad \ldots\ldots \tag{5.2.}$$

$$= -\frac{1}{T}\int_o^T \int_R \mu(t)\rho[u(t)]\,p(x,t) \ln p(x,t)\, dxdt + \frac{1}{T}\int_o^T \mu(t)\rho[u(t)]u(t)H(X';t)dt + K \ln T \quad \square \tag{5.3.}$$

where $\mu(t)$ is the structural parameter which is involved by the equation (4.2.).

As a matter of fact, we could have stated this result in the form of a definition, but we rather use the term of proposition to emphasize that it is a direct consequence of the randomization technique which allowed us to obtain a unified framework for descrete entropy and continuous entropy.

In a like manner, the Minkowskian observation of the pair (H(X, Y;t), H(X;Y;t)) provides the relative or subjective entropy

$$H_r(F_{xy};0,T) := \frac{1}{T}\int_o^T u_{xy}(t)p[u_{xy}(t)]\,[H(X,Y;t) + u_{xy}(t)H(X',Y';t)]\,dt + K \ln T \tag{5.4.}$$

where the subscript xy in $u_{xy}(t)$ and $u_{xy}(t)$ emphasizes that the corresponding variables are related to the pari (X,Y).

Comments. We have shown [14] that $u_{xy}(t)$ necessarily satisfies a composition law in the form

$$u_{xy}(t) = [u_x(t) + u_{y/x}(t)\,] / [1 + u_x(t)u_{y/x}(t)\,] \tag{5.5.}$$

In contrast, the law which is followed by $u_{xy}(t)$ is not so easy to be defined, at first glance. This is an open question. Nevertheless, for white observation, one has of course $u_{xy}(t) = 1$.

For a Markovian process under local Markovian observation, one has the equation

$$\mu_{xy}(t)H(X,Y;t) = H(X,Y;t) + \int_{R^2} p(x,y,t) \ln(|G(x,y,t)|^{\frac{1}{2}} (2\pi e))dxdy \quad (5.6.)$$

where $|G(x,y,t)|$ is the absolute value of the determinant of the transition covariance matrix of the process.

If is clear that while $u_{xy}(t) = u_{yx}(t)$, generally one has $u_{xy}(t) \neq u_{yx}(t)$, so that as a result one has too $H_r(F_{xy};0,T) \neq H_rF_{yx};0,T)$ in most cases. Nevertheless, we can state the following result.

Proposition 5.2. The equality

$$H_r(F_{xy};0,T) = H_r(F_{yx};0,T)$$

holds when and only when the following conditions are simultaneously satisfied, which are

$$u_{x/y}(t) = u_x(t) \text{ and } u_{y/x})t) = u_y(t) \quad \square \quad (5.7.)$$

The proof is a direct consequence of equation (7.5.).

5.3. *Conditional Relative Trajectory Entropy*

In Shannon theory, given two random variables X and Y, the conditional entropy H(Y/X) is defined by the expression

$$H(Y/X) := H(X,Y) - H(X) \quad (5.8.)$$

and this approach is mainly supported by the fact that one has the equality H(X,Y) = H(Y,X). At first glance, analogously one could envision a similar technique to define $H_r(F_{y/x};0,T)$, namely $H_r(F_{y/x};0,t) = H_r(F_{xy};0,T) - H_r(F_x;0,T)$, but this approach would be wrong as we have $H_r(F_{xy};0,T) = H_r(F_{yx};0,T)$.

So in order to circumvent this difficulty, we shall rather consider the explicit expression of H(Y/X). that is to say

$$H(Y/X) := \sum_i p(x_i) H(Y/x_i) \quad (5.9.)$$

and we sahll generalize it in the following way.

Definition 5.1. Let $u_{y/x}(t)$ be defined by Equation (5.5.), then the corresponding *conditional relative trajectory entropy* is defined by the expression.

$$H_r(F_{y/x};0,T) := \frac{1}{T} \int_0^T u_{y/x}(t)\rho\,[u_{y/x}(t)]\,[H(Y/X,X';t) + u_{y/x}(t)H(Y'/X,X';t)\,]\,dt + K \ln T \quad \square \quad (5.10.)$$

Comments. It may be interesting to compare $H_r(F_{y/x};0,T)$ with $H(F_{y/x};0,T)$. To this end, and as a special illustrative case, assume that $u_{y/x}(t) = v(t)$ is small with respect to the unity. The after some calculations, we obtain

$$H(F_{y/x};0,T) - H_r(F_{y/x};0,T) \cong \frac{1}{T} \int_0^T u_{y/x}(t)[H(X'/X;t) - H(X'/X,Y;t) - v(t)H(Y'/X,X';t)\,]\,dt$$

This being so, assume that $u_{y/x}(t) > 0$, what is quite meaningful on a practical standpoint ; and assume further that $v(t) < 0$; we may then have $H(F_{y/x};0,T) > H_r(F_{y/x};0,T)$. Next, suppose that $H(X'/X;t) = 0$ and consequently that $H(X'/X,Y;t) = 0$; when $v(t) > 0$ one may have $H_r(F_{y/x};0,T) > H(F_{y/x};0,T)$.

6. Relative information for stochastic continuous trajectory

6.1. *The Local Approach*

We now have at hand all the prerequisites which will allow us to measure the amount of relative transinformation between two stochastic trajectories X(t) and Y(t) over the time range (0,T) and we state.

Definition 6.1. In the framework (syntax, semantics) and on assuming that the local entreopies are observed via a Minkowskian observation process, the total amount $I(F_y/F_x;0,T)$ of relative (or relativistic) information provided by $(F_x;0,T)$ about $(F_y;0,T)$ is defined by the expression

$$I_r(F_y/F_x;0,T) := H(F_y;0,T) - H_r(F_y/F_x;0,T) \quad (6.1.)$$

$$=: \frac{1}{T}\int_0^T \Big\{ H(Y;t) - u_{y/x}(t)\,p\,[u_{y/x}(t)][H(Y/X,X';t) + u_{y/x}(t)H(Y'/X,X';t)] \Big\}\,dt \qquad (6.2.)$$

$$=: \frac{1}{T}\int_0^T I_r(Y/X;t)\,dt \qquad (6.3.)$$

where $I_r(Y/X;t)$ is the relative information provided by X about Y at the instant t. □

Comments. (i) On the surface, equation, looks like a mere straightforward formal generalization, a bit in the same way as the definition of continuous entropu via formal inference from discrete entropy, but this is a semblance only. Indeed, all this derivation is firmly supported by the randomization technique with respect to time, which is equivalent to assume that time itself involves its own amount of uncertainty.

(ii) According to the first integral mean value theorem, one has

$$I_r(F_y/F_x;0,T) = I_r(Y/X;t_c)$$

for some t_c such that $0 < tc < T$. In other words, the transinformation of the trajectory over the time range (0,T) would be summarized in the instantaneouis transinformation at the instant t_c. We shall see that this remark is of paramount importance when t is no longer time but rather is a space parameter.

(iii) Obviously, when $u_{y/x}(t) = 0$ for every t, one finds again the direct generalization of Shannon transinformartion.

6.2. *The Global Approach*

In subsection 6.1, basically we assumed that the subjectivity of the observer applies to the amount of transinformation at each instant, and that then all these measurements are averaged in the form of an integral with respect to time.

Another valuable approach which merely pictures a different level of observation, is to assume that this subjectivity applies to the trajectory entropies themselves. In other words, these trajectories are themselves subject to the Minkowskian observation process.

In such a case, we shall denote by $U_{y/x}(0,T)$ the conditional subjectivity about the trajectory $(F_y;0,T)$ given $(F_x;0,T)$, and the relative information $I_{rg}(F_y/F_x;0,T)$ so obtained is given by the expression.

$$I_{rg}(F_y/F_x;0,T) := H(F_y;0,T) - \rho\,[U_{y/x}(0,T)]\,[H(F_y/F_x,F'_x,F'_x;0,T) + U_{y/x}(0,T)\,H(F'_y/F_x,F_x,F'_x{}'0,T)] \qquad (6.4.)$$

with the notations

$$H(F_y/F_x,F'_x;0,T) := \frac{1}{T}\int_0^T H(Y/X,X',t)\,dt + \ln T \qquad (6.5.)$$

$$H(F'/F_x,F';0,T) := \frac{1}{T}\int_0^T H(Y'/X,X',t)\,dt + \ln T \qquad (6.6.)$$

Comments. In most general cases $U_{x/y}(t_i,t_f)$ should depend upon the initial instant and the terminal one ; and this dependence upon t_f is quite relevant to describe learning processes, for instance.

(ii) Remark that expression (8.4) is quite consistent with the first integral mean value theorem.

(iii) Which of the transinformation I_r or I_{rg} is the best for applications ? Assume that we are comparing hand written B's and 8's. If we first look at B, and then we look at 8, and then compare the two patterns so obtained, we are then using I_r . But in contrast, if we simultaneously compare B and 8 by using a scanning procedure, the we explicitly refer to I_r.

7. Entropy of deterministic maps

7.1. *Trajectory Entropy of Maps With Respect to a Class of Probability Density Functions*

Definition 7.1. Assume that $\underline{X} \in R^n$ is a random vector and let $\underline{P}$ denote the set of its admissible probability density functions p(x). Then as a direct consequence of the Shannonian entropy, the entropy $H[\underline{f}(.)/\underline{P}]$ of the $R^n \to R^n$ map $\underline{f}(.)$ relatively to $\underline{P}$ is measured by the quantity.

$$H[\underline{f}(.)/\underline{P}] = \max_{p(x)\in \underline{P}} \int_{R^n} p(\underline{x}) \ln |J(\underline{x})|\, d\underline{x} \qquad (7.1.)$$

where $|J(\underline{x})|$ is the Jacobian determinant [2] of $\underline{f}(.)$. □

2. Des entropies faisant intervenir ln $|J(x)|$ ont été considérées, dans le cas général, par J. Fronteau (L'entropie et la physique moderne, CERN, MPS/*Int.* MU/EP, 1966 ; A propos de diverses dynamiques non-hamiltoniennes, *Annales de la Fondation Louis de Broglie*, vol. 1, no 4, p. 179, 1976) et dans le cas linéaire par R. Vallée (Aspect informationnel du problème de la prévision dans le cas d'une observation initiale imparfaite, *Economie Appliquée*, tome 32, n. 2-3, 1979 ; Evolution of a dynamical linear system with random initial conditions, in *Cybernetics and Systems Research,* R. Trappl. (ed.), North Holland Publishing Company, 1982) (N.d.C.R.).

This definition is supported by the equation

$$H[\underline{f}(\underline{X})] = H(\underline{X}) + \int_{R^n} p(\underline{x}) \ln |J(\underline{x})| dx \qquad (7.2.)$$

where the integral is then thought of as the conditional entropy H[f(.)/X].

7.2. *Trajectory Entropy of Degree d of Continuous Maps*

As a special of class P, assume that p(x) is defined by the equation

$$H(\underline{X}) = C \qquad (7.3.)$$

where C is a given constant, then by using Lagarnge parameters for instance, the definition (7.1.) yields the trajectory entropy in the form

$$H_d[f(.)] = \frac{\int_{R^n} |J(\underline{x})|^d \ln |J(\underline{x})| \, d\underline{x}}{\int_{R^n} |J(\underline{x})|^d d\underline{x}} \qquad (7.4.)$$

where $d \in R$ is a parameter which depends upon C : d = d(C).;

Definition 7.2. We shall refer to $H_d[f(.)]$ as to the *trajectory entropy of degree d of* $\underline{f}$(.).

7.3. *Thermodynamic Entropy of Maps and Liapunov Exponent*

For a one-dimensional map f(.) defined on the finite interval [a,b], the trajectory entropy of order zero

$$H_o[f(.).;a,b] = \frac{1}{b-a} \int_a^b \ln |f'(x)| dx \qquad (7.5.)$$

is merely the so-called Liapunov exponent of f(.).

When f(x) is the probability density p(x) of a scalar variable, then one has

$$H_1[p(.)] = -H(x) \qquad (7.6)$$

and this result suggest to consider $H_1[f(.)]$ as being the thermodynamic entropy of f(.).

7.4. *Application to Dynamic Systems*

Consider the dynamic system

$$\dot{x}(t) = -V_x(x) \; ; \; x(0) = x_o, \; x \in R \qquad (7.7.)$$

where V(x) denotes the potential function[3] of the dynamics and $V_x(x)$ holds for the derivative dV(x)/dx. According to the equations (7.3.) and (7.6.) the thermodynamic entropy $H_l[x(.)]$ of the trajectory x(t) on the time range (t_o,t_l) is

$$H_1[x(.);a,b] = \frac{\int_a^b \sigma(x) \ln |V_x(x)| dx}{\int_a^b \sigma(x) \, dx} \qquad (7.8.)$$

with the notations

$$\sigma(x) := -\operatorname{sgn} V_x(x) \; ; \; a := x(t_o) \; ; \; b := x(t_l) \qquad (7.9.)$$

Assume that $\sigma(x) \geqslant 0$ for $a \leqslant x \leqslant, b$, then one has

$$H_l[x(.)] = \frac{1}{b-a} \int_a^b \ln |V_x(x)| dx \qquad (7.10.)$$

In other words, the thermodynamic entropy of th trajectory x(t) is equal to the Liapunov exponent of the potential function V(x).

8. Relative Entropy of Deterministic Maps

Let $Y \in R$ denote a random variable the relative entropy of which is defined by the equation

$$H_r(Y) = \int_R q(y) \rho [v(y)] [-\ln q(y) + v(y)\bar{H}(y)] dy \qquad (8.1)$$

where v(y), $-1 \leqslant v(y) \leqslant +1$ is a function which can be considered as characterizing the subjectivity of the observer, and when $\bar{H}(y)$ is the semantic entropy associated with y.

This being so, we consider the random variable X defined by the transformation Y = f(X) where f(.). is continuously differentiable, and we bear in mind that the probability density p(x) of X is p(x) = q[f(x)] |f'(x) |. We make the following assumptions.

(A1) We suppose that the function u(x) of the relative observation of X is defined by the equation u(x) = v[f(x)].

(A2) We suppose that the semantic functions $\bar{H}(x)$ and $\bar{H}(y)$ are the same.

With these assumptions, we make the transformation Y = f(X) in the equation (8.1.) to obtain

3. Au sens de la théorie des systèmes dynamiques (N.d.C.R.).

$$H_r(y) = \int_R p(x)\rho(u)[-\ln p(x) + \ln |f'| + u(x)\, H[f(x)]]\, dx \qquad (8.2.)$$

$$H_r(X) + \int_R p(x)\rho(u)\ [\ln |f'(x)| + u(x)[\bar{H}[f(x) - \bar{H}(x)]]dx \qquad (8.3)$$

By using an approach similar to the derivation outlined in Subsection 7.1. but applied to Equation (8.3.), we then obtain the.

Definition 8.1. The relative (or subjective) entropy of the map f(.). with respect to the class P̲ of probability density functions defined by Equation (7.3.) is

$$H_{rd}[f(.)] = \frac{\int_a^b [\ |f'(x)|^d p(u) \ln |f'(x)| + u(x)[\bar{H}(f) - \bar{H}(x)]]dx}{\int_a^b |f'(x)|^d\ dx} \qquad (8.4)$$

9. Conclusions

Our thesis is that the information theory as initiated by Shannon a few decades ago contains in itsefl the seeds for its self-generalization. This is true to settle the apparent discrepancy there is between discrete entropy and continuous entropy, this is true to define the entropy of deterministic and stochastic trajectories ; and this is true to introduce syntax and semantics in the framework. In this way, *one can claims that Shannon theory is a self-referencial theory.*

As concluding remardks, we shall summarize the potential of the approach outlined above.

(i) We considered linear observation only, but it is clear that the principle of informational invariance applies also to non linear observation.

(ii) The entropy of deterministic patterns, as so derived, is a by-product of Shannon theory, therefore a unity which will be of help to expand a thermodynamics of forms, for instance by using some ideas of Mendès France et alii [4](. Moreover it is not impossible that our new definition of deterministic entropy refines the results obtained by this author on the temperature of curves.

(iii) The theory applies to discrete phenomena by using our concept of «total entropy of a discrete variable» and by this way we should be able to exhibit its relations with the theory of fractals.

(iv) At first glance, it seems that the approach can be generalized to the multi-dimensional case without too much difficulty.

(v) The equation derived in subsection (7.3.) is interesting as it relates Liapunov exponent and entropy of order one, and in this way it would not be surprizing that the latter be a better measure of how chaotic is a dynamics than the former.

(vi) Given this remark in subsection (7.3.), can we classify potential functions by their Liapunov exponents and then have a new look at the catastrophe theory ?

References

[1] ACZEL J., DAROCZY Z., A mixed theory of information I. Symmetric, recursive and measurable entropies of randomized systems of events ; *RAIRO Theoretical, Computer Science,* Vol. 12, No 2, pp 149-155, 1978.

[2] ASHBY W.R., *An Introduction to Cybernetics,* Chapman and Hall, London, 1956.

[3] BRILLOUIN L., *Science and Information Theory*, Academic Press, New York, 1956.

[4] DUPAIN Y., KAMAE T., MENDES FRANCE, Can one measure the temperature of a Curve ? *Archive for Rational Mechanics and Analysis,* Vol. 94, No 2, pp 155-163, 1986.

[5] GALLAGER R.G., *Information Theory and Reliable Communication,* John Wiley, New York, 1968.

[6] HARTLEY R.V., Transmission of information, *The Bell System Technical Journal*, Vol 7, pp 535-563, 1928.

[7] HIGASHI H., KLIR G.J., Measures of uncertainty and information based on possibility distributions, *Int. Journal of General Systems,* Vol 9, pp 43-58, 1982.

[8] JUMARIE G., Further advances on the general thermodynamics of open systems via information theory, Effective entropy, negative information, *Int. Journal of Systems Science*, Vol 6, No 3, pp 249-268, 1975.

[9] JUMARIE G., Théorie relativiste de l'information I, II, III, IV, V, *Annales des Télécommunications*, Vol 33, Nos 1-2, pp 13-27, 1978 ; Vol 34, Nos 9-10, pp. 491-507, 1979 ; Vol 34, Nos 11-12, pp 521-530, 1979 ; Vol 35, Nos 7-8, pp 281-296, 1980 ; Vol 37, Nos 5-6, pp 201-212, 1982.

[10] JUMARIE G., Définition et application d'une mesure de l'infor-

mation tenant compte de la signification des symboles, *Annales des Télécommunications*, Vol 39, Nos 11-12, pp 523-537, 1984.

[11] JUMARIE G., *Subjectivity, Information, Systems. Introduction to a Theory of Relativistic Cybernetics*, Gordon and Breach, London, New York, 1986.

[12] JUMARIE G., Some approaches, to the measure of the amount of information involved by a form, *Sujet. Analysis, Modelling, Simulation*, Vol 3, No 6, pp 479-506, 1986.

[13] JUMARIE G., A Minkowskian theory of observation. Application to uncertainty and fuzziness, *Fuzzy Sets and Systems* (to appear), Vol 24, No 1, 1987.

[14] JUMARIE G., Analysis of nonlinear stochastic distributed systems by using the dynamic equations of their state moments, *Proceedings of the 1987 American Control Conference*, IEEE Publications, New York, 1987.

[15] JUMARIE G., *Relative Information. Theories and Applications*, Springer Verlag, Berlin, New York (to appear), 1987-1988.

[16] MANDELBROT B.B., *The Fractal Geometry of Nature*, W.H. Freeman and Company, New York, 1977, 1982, 1983.

[17] NYQUIST H., Certain topics in telegraph transmission theory, *AIEE Trans.*, Vol 47, 9617, 1928.

[18] SHANNON C.E., A mathematical theory of communication I, II, *The Bell System Technical Journal*, Vol 27, pp 379-423, 623-656, 1948.

REVUE INTERNATIONALE DE SYSTEMIQUE
Vol. 2, N° 1, 1988, pp. 47 à 62

HASARD ET SYSTEMES, QUELQUES REMARQUES

Jacques BONITZER

Professeur Honoraire [1]

Ecole Nationale des Ponts et Chaussées

Résumé
Le renouvellement de la théorie des Probabilités au XXe siècle, et les problèmes des Systèmes, appellent un renouvellement correspondant de l'épistémologie du hasard. L'article insiste particulièrement sur l'existence de structures spécifiques sous-jacentes à tout phénomène aléatoire, sur le rôle dominant et les implications des propriétés d'additivité en théorie des Probabilités, et sur la nécessité de comprendre le sujet soumis à des effets aléatoires comme sujet social.

Abstract
Renewal of Probability Theory in the XXth Century, as well of Systems problems, call for a renewal of Chance epistemology as well. The author particuliary emphasizes following items : existence of specific structures underlying every chance phenomenon; ruling character and implications of additivity properties in Probability Theory ; necessity of considering subjects submitted to random effects with their social character.

Les modèles systémiques d'émergence du sens et d'organisation par le bruit font appel aux techniques du calcul des Probabilités et de la théorie de l'Information à un niveau tel, et de telle façon, qu'ils invitent à renouveler la réflexion épistémologique sur le concept même de hasard. Leur invite, au demeurant, ne fait que renforcer celle qu'implique le renouvellement en profondeur de la théorie des Probabilités

1. 8, rue Degas, F. 75016, Paris, France.